Nonparametric adaptive estimation for discretely observed Lévy processes

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von

Dipl.-Math. Julia Johanna Kappus

Präsident der Humboldt-Universität zu Berlin :
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II :
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Markus Reiß
2. Prof. Dr. Michael Neumann
3. Prof. Dr. Fabienne Comte

eingereicht am: 23.05.2012
Tag der Verteidigung: 28.09.2012

Abstract

This thesis deals with nonparametric estimation methods for discretely observed Lévy processes.

The following statistical framework is considered: A Lévy process X having finite variation on compact sets and finite second moments is observed at low frequency. In this situation, the jump dynamics is fully described by the finite signed measure $\mu(dx) = x\nu(dy)$. The goal is to estimate, nonparametrically, some linear functional of μ.

In the first part of this thesis, kernel estimators are constructed and upper bounds on the corresponding risk are provided. From this, rates of convergence are derived, under global as well as under local regularity assumptions on the Lévy measure. For particular cases, minimax lower bounds are proved. The rates of convergence are thus shown to be optimal in the minimax sense.

The focus of this thesis lies on the problem of adaptive estimation, more precisely, on the data driven choice of the smoothing parameter, which is being considered in the second part.

Since nonparametric estimation methods for Lévy processes have strong structural similarities with with nonparametric density deconvolution with unknown error density, both fields are discussed in parallel and the concepts are developed in generality, for Lévy processes as well as for density deconvolution.

The choice of the bandwidth is realized, using techniques of model selection via penalization.

The principle of model selection via penalization usually relies on the fact that the fluctuation of certain stochastic quantities can be controlled by penalizing with a deterministic term. Contrarily to this, the variance is unknown in the setting investigated here and the penalty term is hence itself a stochastic quantity.

It is the main concern of this thesis to develop strategies to dealing with the stochastic penalty term. The most important step in this direction will be a modified estimator of the unknown characteristic function in the denominator, which allows to make the pointwise control of this object uniform on the real line.

The main technical tools involved in the arguments are concentration inequalities of Talagrand type for empirical processes.

Contents

Introduction

Lévy processes, continuous time stochastic processes with independent and stationary increments, are the building blocks for a large number of continuous time models with jumps which play an important role, for example, in the modelling of financial data. Let us mention exponential Lévy models, see e.g. Carr et al. [12, 14], hyperbolic Lévy motions, see Eberlein and Keller [24, 40], time changed Lévy processes, see Carr et al. [15], or stochastic volatility models, see Carr et al. [13]. For an overview of the relevance of Lévy processes in financial applications, we refer to the monograph by Cont and Tankov [22].

For this reason, statistical inference for Lévy processes has received considerable attention during the past decade and is both, a topic of great theoretical relevance and also an important issue for practitioners.

A Lévy process X is determined by three parameters, the so-called characteristic triplet: The volatility $\sigma^2 \geq 0$, the drift parameter $\gamma \in \mathbb{R}$ and the Lévy measure or jump measure ν.

We consider, in the present thesis, the following statistical framework: Given low frequency observations

$$X_\Delta, \cdots, X_{2n\Delta}$$

of a Lévy process $X = \{X_t : t \geq 0\}$ having finite variation on compact sets and finite second moments, we consider the problem of estimating, nonparametrically, the finite signed measure $\mu(\,\mathrm{d}x) := x\nu(\,\mathrm{d}x)$. More precisely, our goal is to estimate a linear functional of μ.

Background and related work

Parametric estimation methods for Lévy processes are already quite well studied. We refer, at this point, to the results and discussion presented in [72]. However, the application of standard parametric methods to Lévy models raises a number of difficulties. This is due to the fact that for parsimonious Lévy densities, the marginal distributions are often analytically intractable or not even expressible in a closed form. Moreover, there is always the danger of model-misspecification, since the Lévy measure is an infinite dimensional object and a parametric model may easily turn out to be far from the truth.

For this reason, nonparametric estimation methods for Lévy processes have become more and more popular during the past few years. The work by Belomestny and Reiß in 2006 [4] dealing with exponential Lévy models, can be seen as a starting point for the recent progress in the theory. For related work, see Belomestny [3], Trabs [68] and Söhl [64].

Nonparametric estimation for continuously and high frequently observed Lévy processes has since then been considered in a number of articles by Comte and

Genon-Catalot [17, 19], by Figueroa-López, see [28, 29] and subsequent papers, and Figueroa-López and Houdré [30, 27].

Indeed, the high frequency approach is not only challenging from a theoretical point of view, but also of great practical relevance since in the field of financial applications, there are typically high frequency data available.

In the present work, we do not deal with a high frequency model, but investigate estimation methods for Lévy processes observed at low frequency. That is, we do not consider a double asymptotic framework where the observation distance Δ tends to zero while the time horizon T increases to infinity, but place ourselves in the situation where Δ is fixed.

The theory of nonparametric estimation for Lévy processes observed at low frequency has been initiated by Neumann and Reiß in 2009, see [56]. The particular case of nonparametric estimation for processes with finite jump activity has been considered by Gugushvili [33, 34] and the problem of adaptive nonparametric estimation for low frequency observed Lévy processes has first been studied by Comte and Genon-Catalot, see [18]. Recently Nickl and Reiß [58] have proved a Donsker theorem for the distribution function of the Lévy measure for certain classes of Lévy processes.

It is interesting to note that the continuous time setting has many common points with a density estimation problem, which is a classical and widely studied topic in nonparametric statistics. Just alike one applies a smoothing procedure to the empirical distribution to obtain an estimate of the underlying distributional density, one uses the empirical jump measure as an estimator of the true underlying jump measure and applies some smoothing procedure in order to infer the underlying jump density. This is the framework which is being considered in [30].

This analogy carries over to the problem of estimating the Lévy measure when high frequency observations of the underlying process are available. In this setting, one can discretise the procedure and make use of the fact the jumps are not directly feasible, but observable in the limit. See, for example, [17, 19] and [29].

Striking similarities of nonparametric estimation problems for Lévy processes with classical and well-studied topics of nonparametric statistics are also found in the low frequency framework which we consider in this thesis. Our statistical model can be seen to belong to the broader class of statistical inverse problems. We face here a deconvolution problem which is intimately connected to classical density deconvolution. To mention only a few of the various papers on this subject, we refer to the work by Carroll and Hall [16], Stefanski [65], Fan [26], Efromovich [25], Comte et al. [21, 9] and Meister [54].

More precisely, the problem at hand can be compared to a density deconvolution problem with unknown distribution of the errors. It is interesting to note that this topic is still an area of very recent and ongoing research. For earlier work on the subject, let us mention Neumann [57, 55], Johannes [36] and Comte and Lacour [20].

The fact that our area of research has various common points with the classical model of density deconvolution motivates us to develop our ideas not only for nonparametric estimation for Lévy processes, but also have a closer look at density deconvolution and discuss some of the concepts in parallel.

Main results

For a Lévy processes X having finite variation on compact sets and finite first moments $\mu(\,\mathrm{d}x) = x\nu(\,\mathrm{d}x)$ is a finite signed measure which completely describes the jump dynamics.

Given discrete time observations of X, with observation distance Δ kept fixed, our goal is to estimate, nonparametrically, some linear functional of μ.

The problem of estimating the density of μ with L^2-loss has been treated in [18]. For processes having finite jump activity, the estimation of the jump density with L^2-loss has been considered in [33] and [34]. The estimation of integrals of the form $\int f(x)\mu(\,\mathrm{d}x)$ for a finite measure μ connected to ν and for smooth test functions f has been investigated in in [56]. To our knowledge, the estimation of general linear functionals of μ in presence of low frequency data, covering the particularly interesting cases of point estimation and estimating derivatives (if a smooth Lévy density exists) as well as the estimation of integrals over compact sets, which is being considered here, has not been treated before in the literature.

We construct kernel estimators for linear functionals of μ and provide upper bounds on the corresponding risk. From this, we derive rates of convergence under global regularity assumptions on μ, measured in a Sobolev sense, as well as under local regularity assumptions, measured in a Hölder sense.

We provide minimax lower bounds for point estimation and estimating integrals and find that our rates of convergence are minimax optimal.

We also have a closer look at the related field of density deconvolution. Even though this class of problems has various common points with the Lévy model investigated here, the attempt to generalise the concepts from density deconvolution to the Lévy model raises certain difficulties which are due to the special structural properties of Lévy measures. We will argue that the point of view of measuring the performance of estimators on a global Sobolev scale, which is considered, for example, by Butucea and Comte [9] in the density deconvolution model, is often inappropriate in the Lévy model. This reasoning relies on an analysis of the dependence between the jump activity a Lévy process and the smoothness of the corresponding distributional density, which is presented in Chapter 1.

The second part of this thesis is devoted to the problem of adaptive estimation, that is, to the data driven choice of the smoothing parameter. This is done by applying techniques of model selection via penalization.

The model selection approach in its present form has been initiated by Birgé and Massart in a series of papers in the late 90s and early 00s, see for example [6, 5] and [53].

The application to density deconvolution problems has been investigated by Comte et al [21] and by Comte and Butucea [9] and Comte and Lacour [20]. The estimation of linear functionals using model selection techniques in the Gaussian white noise model has been treated by Laurent, Ludeña and Prieur, see [44].

Adaptive estimation of linear functionals using approaches different from model selection, typically in the spirit of Lepski's method, has been considered, since the early 90s, by Lepski [46, 47, 48], Lepski and Spokoiny [49], by Tsybakov [69], Tsybakov and Klemelä and by Cai and Low [10, 11].

One might suspect that the only thing which is left to be done here is to generalise some of the classical and well known concepts from the white noise and density deconvolution framework to the Lévy model.

However, the setting which is being considered here raises certain difficulties which are in fact non-standard and the treatment is not obvious. This is due to the fact that we face here a problem of model selection with unknown variance.

Adaptive estimation via penalization usually relies on the fact that one can control the fluctuation of certain stochastic quantities by penalizing with a deterministic correction term. To understand the ideas behind this concept, we refer to Birgé [5].

In density deconvolution models, the correction term involves the characteristic function of the noise, and in the Lévy model, the characteristic function of X_Δ. Now, this object is clearly not feasible so the correction term is itself a stochastic quantity.

We propose a new approach to the problem of model selection with unknown variance and discuss the application to the Lévy model as well as to deconvolution with unknown error distribution. The main technical step will be an extension of a classical result by Neumann, making the pointwise control on an estimator of the characteristic function appearing in the denominator uniform on the real line.

In the framework of estimating the density of μ with L^2-loss, which is closely connected to the model which is being considered here, the problem of the unknown variance has been treated, using an additional a priori assumption on the size of the model, see [18]. However, since this assumption involves itself some prior knowledge of the characteristic function, it is certainly very restrictive. In the approach which is considered here, we can completely drop this assumption, so our reasoning no longer relies on the prior information on the unknown object.

Compared to the recent work by Comte and Lacour [20], where density deconvolution with unknown distribution of the errors is considered, the main difference of our approach is the fact that we consider the problem of the stochastic penalty term from a different point of view. This will allow a more general treatment, since we can drop certain semiparametric assumptions on the characteristic function of the noise, and might provide some new insight into the nature of the problem.

This thesis is organised as follows: In Chapter 1, we give a concise overview of the basic facts and definitions concerning Lévy processes and infinitely divisible distributions and give a brief discussion on smoothness properties of infinitely divisible laws.

In Chapter 2, we construct kernel estimators of linear functionals for Lévy processes observed at low frequency and for density deconvolution with unknown error distribution. We derive upper risk bounds and rates of convergence under regularity assumptions and prove that the rates of convergence thus obtained are minimax optimal.

In Chapter 3, we consider the problem of the data driven choice of the smoothing parameter, using methods of model selection via penalization, and propose a new approach to the problem of model selection with unknown variance.

Acknowledgements

I thank Markus Reiß for his guidance and for useful hints and inspiring discussions. I also wish to thank Fabienne Comte and Valentine Genon-Catalot for giving me the opportunity for a research stay at Université Paris Déscartes and for teaching me about model selection.

I thank the Deutsche Akademische Austauschdienst for supporting my stay in Paris.

I also want to express my thankfulness to the present and former members of my research group, Markus Fischer and Markus Bibinger, Hartmuth Henkel and Hilmar Mai, Jakob Söhl and Matthias Trabs, for working together in a pleasant atmosphere, and to Andrea Acker, Christine Grün and Emeline Schmisser.

Most of all, I thank my family for so much support and help and understanding.

Chapter 1

Lévy processes and infinitely divisible distributions

1.1 Basic facts and definitions

In the present section, we briefly summarise the most important facts and definitions concerning Lévy processes and the related concept of infinitely divisible distributions, which are used throughout the rest of this thesis. All results collected here can be found in the literature on Lévy processes. For the proofs and for further reading, we refer to the textbooks by Sato [63] and Applebaum [2] and to Cont and Tankov [22].

Definition 1.1.1. *A continuous time stochastic process $X = \{X_t : t \in \mathbb{R}^+\}$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ taking values in $\mathbb{R}^d$ is called a* Lévy process *if the following conditions are satisfied:*

(i) $X_0 = 0$ *a.s.*

(ii) X *has* independent increments*: For arbitrary $n \in \mathbb{N}$ and time points $0 = t_0 < t_1 < \cdots < t_n$, the random variables $X_{t_1} - X_{t_0}, \cdots, X_{t_n} - X_{t_{n-1}}$ are independent.*

(iii) X *has* stationary increments*: For arbitrary $s, t \geq 0$, the distribution of $X_{t+s} - X_t$ does not depend on t.*

(iv) X *is* stochastically continuous*, that is, for arbitrary $t \geq 0$,*

$$X_s \xrightarrow{\mathbb{P}} X_t, \ \ s \to t.$$

(v) X *has almost surely* càdlàg paths*, that is, there is some measurable set $\Omega_0 \in \mathcal{A}$ such that $\mathbb{P}(\Omega_0) = 1$ and for all $\omega \in \Omega_0$, $t \mapsto X_t(\omega)$ is right continuous with left hand limits.*

The following are the most elementary and, at the same time, the most important examples of Lévy processes:

Examples 1.1.2.

(i) A Brownian motion with linear drift is, by definition, a Lévy process with continuous sample paths.

(ii) *A Poisson process is a Lévy process. The same is true for compound Poisson processes, that is, Poisson processes where the sizes of the jumps are random and independent of the jump times.*

Indeed, Brownian motion and compound Poisson processes are the building blocks for the whole class of Lévy processes. Any Lévy process X can be seen as the sum $X = X_1 + X_2 + X_3$ of three independent components, where X_1 is a Brownian motion with linear drift, X_2 is a compound Poisson process which collects the jumps of size larger than 1 and X_3 is a square integrable martingale, which is found as the limit of infinitely many compensated compound Poisson processes. For the exact formulation and proof of this so-called *Lévy-Itô decomposition*, see e.g. Chapter 4 in Sato [63].

Lévy processes are intimately connected to the concept of infinitely divisible distributions which is made precise in the following definition:

Definition 1.1.3. *A probability measure $\mathbb{P}$ on $\mathbb{R}^d$ is* infinitely divisible *if for any positive integer n, there is a probability measure $\mathbb{P}_n$ such that $\mathbb{P}$ is the n-fold convolution of $\mathbb{P}_n$ with itself,*

$$\mathbb{P} = (\mathbb{P}_n)^{*n} := \underbrace{\mathbb{P}_n * \cdots * \mathbb{P}_n}_{n\text{-times}}. \tag{1.1.1}$$

A random variable Z taking values in $\mathbb{R}^d$ is called infinitely divisible if $\mathbb{P}^Z$ is infinitely divisible. This is equivalent to stating that there are i.i.d. random variables $Z_1, \cdots, Z_n$ such that

$$Z = Z_1 + \cdots + Z_n. \tag{1.1.2}$$

Lévy processes and infinitely divisible distributions are connected as follows:

Theorem 1.1.4.

(i) *If $X = \{X_t : t \in \mathbb{R}^+\}$ is a Lévy process, then for any $t \geq 0$, $\mathbb{P}^{X_t}$ is infinitely divisible and for any positive integer n, $\mathbb{P}^{X_t} = \left(\mathbb{P}^{X_{\frac{t}{n}}}\right)^{*n}$.*

(ii) *Conversely, if $\mathbb{P}$ is an infinitely divisible law on $\mathbb{R}^d$ and $t > 0$, then there is a Lévy process X such that $\mathbb{P}^{X_t} = \mathbb{P}$.*

Examples 1.1.5. *Let $d = 1$.*

(i) *For $\beta, \lambda > 0$, the $\Gamma(\beta, \lambda)$-distribution has Lebesgue density*

$$f_{\beta,\lambda}(x) = \frac{\lambda^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\lambda x} 1_{(0,\infty)}(x) \tag{1.1.3}$$

and characteristic function

$$\varphi_{\beta,\lambda}(u) = \left(1 - \frac{iu}{\lambda}\right)^{-\beta}. \tag{1.1.4}$$

We do thus have for arbitrary $n \in \mathbb{N}$:

$$\varphi_{\beta,\lambda}(u) = \left(\left(1 - \frac{i}{\lambda}u\right)^{-\frac{\beta}{n}}\right)^n = \left(\varphi_{\frac{\beta}{n},\lambda}(u)\right)^n. \tag{1.1.5}$$

Since the convolution operation is equivalent to pointwise multiplication of characteristic functions, this implies readily that the $\Gamma(\beta,\lambda)$-distribution is infinitely divisible.

A Lévy process taking values in $\mathbb{R}$ is called a Gamma process *with parameters β and λ if X_1 has a $\Gamma(\beta,\lambda)$-distribution.*

(ii) *For $\alpha \in (0,2]$, an α-*stable distribution *with parameters $\beta \in [-1,1]$, $c > 0$ and $\gamma \in \mathbb{R}$ has characteristic function*

$$\varphi_{\alpha,\beta,c,\tau}(u) = \begin{cases} \exp\left(-c|u|^\alpha \left(1 - i\beta \tan\frac{\pi\alpha}{2} \operatorname{sgn} u\right) + i\gamma u\right) & \text{if } \alpha \neq 1 \\ \exp\left(-c|u|^\alpha \left(1 + i\beta\frac{2}{\pi}(\operatorname{sgn} u)\log|u|\right) + i\gamma u\right) & \text{if } \alpha = 1 \end{cases}. \tag{1.1.6}$$

Special cases are the normal distribution ($\alpha = 2$), Cauchy distribution ($\alpha = 1$, $\beta = 0$) and Lévy distribution ($\alpha = 1/2$, $\beta = 1$). In most cases, the probability densities of stable distributions are not expressible in a closed form.

For arbitrary $n \in \mathbb{N}$ we have

$$\varphi_{\alpha,\beta,c,\gamma}(u) = \left(\varphi_{\alpha,\beta,\frac{c}{n},\frac{\gamma}{n}}(u)\right)^n, \tag{1.1.7}$$

which implies infinite divisibility of α-stable distributions.

*We call X an α-*stable process *with parameters β, c and γ if X_1 has an α-stable distribution with parameters β, c and γ.*

The following result is known as the *Lévy-Khintchine formula* or *Lévy- Khintchine representation* of infinitely divisible distributions.

Theorem 1.1.6 (Lévy-Khintchine formula)**.**

(i) *Let $\mathbb{P}$ be an infinitely divisible distribution on $\mathbb{R}^d$ and let φ denote its characteristic function. Then there is a symmetric, nonnegative definite matrix $\Sigma \in \mathbb{R}^{d\times d}$, a vector $\gamma \in \mathbb{R}^d$ and a measure ν on $\mathbb{R}^d$ satisfying*

$$\nu(\{0\}) = 0 \text{ and } \int \left(|x|^2 \wedge 1\right) \nu(\,dx) < \infty \tag{1.1.8}$$

such that $\varphi(u) = \exp(\Psi(u))$ with

$$\begin{aligned}\Psi(u) &= -\frac{1}{2}\langle u, Au\rangle + i\langle\gamma, u\rangle \\ &\quad + \int \left(e^{i\langle u,x\rangle} - 1 - i\langle u,x\rangle 1(\{x \leq 1\})\right)\nu(\,dx).\end{aligned} \tag{1.1.9}$$

Σ, γ and ν are uniquely defined.

(ii) *Conversely, given a nonnegative definite matrix* $\Sigma \in \mathbb{R}^{d\times d}$*, a vector* $\gamma \in \mathbb{R}^d$ *and a measure* ν *satisfying (1.1.8), there is an infinitely divisible distribution* $\mathbb{P}$ *with characteristic function* $\varphi(u) = \exp(\Psi(u))$ *such that* $\Psi(u)$ *is given by (1.1.9).*

(iii) *Let* X *be a Lévy process and let* $\varphi_1(u) = \exp(\Psi(u))$ *be the characteristic function of* X_1*. Then* X_Δ *has characteristic function* $\varphi_\Delta(u) = \exp(\Delta\Psi(u))$*.*

Definition 1.1.7. *In the situation of the preceding statement,* $\Psi(u)$ *is called the* characteristic exponent *and the triplet* (Σ, γ, ν) *is called the* characteristic triplet *or* Lévy-Khintchine triplet*.*

The matrix Σ *is called the* Gaussian covariance matrix *and the measure* ν *is called the* Lévy measure *or* jump measure*.*

It was mentioned that a Lévy process X is the sum of a Gaussian part X_1 and a pure jump part X_2 independent of X_1. Indeed, Σ and γ are the covariance matrix and drift of the Brownian component while the Lévy measure ν determines the jump activity of X in the following sense:

Lemma 1.1.8. *Let* X *be a Lévy process on* $\mathbb{R}^d$ *with characteristic triplet* (Σ, γ, ν)*. Then we have for any* $B \in \mathcal{B}(\mathbb{R}^d)$ *that*

$$\nu(B) = \mathbb{E}\left[\#\left\{t \in [0,1] : X_t - X_{t-} \in B \setminus \{0\}\right\}\right], \tag{1.1.10}$$

with $X_{t-} := \lim_{s\uparrow t} X_s$ *denoting the left hand limit, which exists by definition of a Lévy process.*

Lemma 1.1.8 tells us that $\nu(B)$ is the expected number of jumps per time unit with magnitude in B.

It follows from (1.1.8) and (1.1.10) that for any measurable set B bounded away from the origin, X has almost surely only finitely many jumps per time unit with size in B, while there may occur (countably) infinitely many small jumps in a finite time interval.

The next result will tell us that the moments of the Lévy measure determine the moments of the corresponding infinitely divisible law.

Lemma 1.1.9. *Let* $\mathbb{P}$ *be infinitely divisible and let* ν *be the corresponding Lévy measure. Let* $g(x) = |x|^\beta$ *with arbitrary* $\beta > 0$ *or* $g(x) = \exp\left(c|x|^\beta\right)$ *with arbitrary* $c > 0$ *and* $\beta \in (0,1]$*. Then*

$$\int g(x)\,\mathbb{P}(\,dx) < \infty \Leftrightarrow \int_{\{|x|>1\}} g(x)\nu(\,dx) < \infty. \tag{1.1.11}$$

Lemma 1.1.9 immediately implies that finiteness of the g-moments is not a time-dependent property. That means, given a Lévy process X, $\mathbb{E}[g(X_t)]$ is finite for *one* $t > 0$ if and only if $\mathbb{E}[g(X_t)]$ is finite for *any* $t > 0$.

The next result explains the dependence between certain pathwise properties of a Lévy process and the corresponding Lévy measure.

Lemma 1.1.10. *A Lévy process X has almost surely finite variation on compact sets if and only if X has no Gaussian component and the jump activity is moderate in the sense that*

$$\int_{\{|x|\leq 1\}} |x|\nu(\,dx) < \infty. \tag{1.1.12}$$

We will also need the following versions of the Lévy-Khintchine representation:

Lemma 1.1.11. *Let $d = 1$.*

(i) Let X be a Lévy process for which X_1 has a finite second moment. Then $\nu_\sigma(\,dx) := \sigma^2\delta_0(\,dx) + x^2\nu(\,dx)$ is a finite measure and the characteristic exponent can be written in the following form:

$$\Psi(u) = iu\gamma_1 + \int \frac{(e^{iux} - 1 - iux)}{x^2}\nu_\sigma(\,dx), \tag{1.1.13}$$

where the integrand is continuously extended at zero and

$$\gamma_1 = \gamma + \int_{\{|x|>1\}} x\nu(\,dx). \tag{1.1.14}$$

In this parametrisation, γ_1 is called the center *of $\mathbb{P}^{X_1}$ and we have $\mathbb{E}\left[X_1\right] = \gamma_1$.*

(ii) Let X be a Lévy process having finite variation on compact sets and finite first moments. Then $\mu(\,dx) := x\nu(\,dx)$ is a finite signed measure and the characteristic exponent can be written in the following form:

$$\Psi(u) = iu\gamma_2 + \int \frac{(e^{iux} - 1)}{x}\mu(\,dx) \tag{1.1.15}$$

with

$$\gamma_2 = \gamma - \int_{\{|x|\leq 1\}} x\nu(\,dx). \tag{1.1.16}$$

In this representation, the parameter γ_2 is called the drift.

Finally, the following result will be important for the statistical analysis:

Lemma 1.1.12. *Let $\mathbb{P}$ be infinitely divisible. Then the characteristic function φ of $\mathbb{P}$ has no zeros, that is, for arbitrary $u \in \mathbb{R}^d$, $\varphi(u) \neq 0$.*

We conclude this section with some more examples:

Examples 1.1.13.

(i) Let X be a compound Poisson process with intensity λ and jump distribution m. Then the characteristic function of X_1 is

$$\varphi(u) = \exp\left(\lambda \int \left(e^{iux} - 1\right) m(\,dx)\right), \tag{1.1.17}$$

so the Lévy measure is finite and given by

$$\nu(\,dx) = \lambda m(\,dx). \tag{1.1.18}$$

Conversely, every Lévy process having a finite Lévy measure ν is the sum of a Brownian motion with linear drift an independent compound Poisson process with jump distribution $m(\,dx) = \frac{\nu}{\nu(\mathbb{R})}(\,dx)$ and intensity $\lambda = \nu(\mathbb{R})$.

(ii) *Let X be a Gamma process with parameters β and λ. Then the Lévy measure has a Lebesgue density g which reads as follows:*

$$g(x) = \beta x^{-1} e^{-\lambda x} 1_{(0,\infty)}(x). \tag{1.1.19}$$

Lemma 1.1.9 tells us that for arbitrary $c < \lambda$, and any $t > 0$, $\mathbb{E}\left[\exp\left(c|X_t|\right)\right]$ is finite and the polynomial moments of all orders are finite.

Moreover, we can use Lemma 1.1.10 to see that X has finite variation on compact sets.

(iii) *The Lévy measure of an $\alpha-$stable distribution has Lebesgue density*

$$g(x) = c_1 x^{-\alpha-1} 1_{(0,\infty)} + c_2 |x|^{-\alpha-1} 1_{(-\infty,0)} \tag{1.1.20}$$

with constants $c_1 = \frac{\tau}{2}(1+\beta)$ and $c_2 = \frac{\tau}{2}(1-\beta)$. By Lemma 1.1.10, an $\alpha-$ stable Lévy process has finite variation on compact sets if and only if $\alpha < 1$. On the other hand, by Lemma 1.1.9 it has a finite first moment if and only if $\alpha > 1$.

(iv) *Tempered α-stable laws are constructed by multiplying the Lévy measure of an α-stable law with some decreasing exponential. That is, for nonnegative constants λ_+ and λ_-, the Lévy measure of a tempered α-stable distribution has Lebesgue density*

$$g(x) = c_1 e^{-\lambda_+ x} x^{-\alpha-1} 1_{(0,\infty)} + c_2 e^{\lambda_- x} |x|^{-\alpha-1} 1_{(-\infty,0)}. \tag{1.1.21}$$

By construction, a tempered α-stable law and has finite variation on compact sets if and only if $\alpha < 1$ and the polynomial moments of all orders are finite. Moreover, $\mathbb{E}\left[\exp\left(c|X_t|\right)\right] < \infty$ holds true for $c < (\lambda_- \wedge \lambda_+)$.

1.2 Smoothness

It will be important for the statistical analysis to clarify the connection between the jump activity of a Lévy process and the smoothness of the distributional density of the corresponding infinitely divisible law (if a density exists).

For sake of simplicity, we consider, in the present section, the one dimensional case.

In what follows, let $\mathbb{P}$ be an infinitely divisible distribution on $\mathbb{R}$, let φ be its characteristic function and ν the corresponding Lévy measure. Moreover, let

$\widetilde{\nu}(\,\mathrm{d}x) := \frac{1}{2}(\nu(\,\mathrm{d}x) + \nu(\,\mathrm{d}(-x)))$ be the symmetrised version of ν and let

$$G(x) := 2\widetilde{\nu}([x,1])1_{(0,\infty)}(x). \tag{1.2.1}$$

The following result, which has first been proved in 1968 by Orey[59] and can also be found in Chapter 28 in [63] states that a high activity of small jumps, measured in an appropriate sense, implies the existence of a smooth distributional density.

Theorem 1.2.1. *Assume that for some $\alpha \in (0,2)$,*

$$\liminf_{r\downarrow 0} \frac{\int_{-r}^{r} x^2\nu(\,dx)}{r^{2-\alpha}} > 0. \tag{1.2.2}$$

Then there exist positive constants C_φ and c_φ such that

$$\forall u \in \mathbb{R}: \ |\varphi(u)| \leq C_\varphi \exp\left(-c_\varphi |u|^\alpha\right). \tag{1.2.3}$$

This theorem implies, in particular, that $\mathbb{P}$ possesses a Lebesgue density which has derivatives of all orders.

In terms of G, one can show the following statement, which can also be found in [59]:

Theorem 1.2.2. *Assume that there are positive constants α and β such that*

$$\liminf_{x\to 0} \frac{G(x)}{x^{-\alpha}} > 0 \tag{1.2.4}$$

holds and moreover,

$$\limsup_{x\to 0} \frac{G(x)}{x^{-\beta}} < \infty. \tag{1.2.5}$$

Then there are constants C_φ and c_φ such that

$$\forall u \in \mathbb{R}: |\varphi(u)| \leq C_\varphi \exp\left(-c_\varphi |u|^\gamma\right) \tag{1.2.6}$$

holds with

$$\gamma = 2(\alpha-\beta)/\alpha + \beta. \tag{1.2.7}$$

It is important to note that (1.2.5) cannot be omitted and moreover, the condition

$$\int_{-1}^{1} |x|^\alpha \nu(\,\mathrm{d}x) = \infty \tag{1.2.8}$$

which is related to the condition (1.2.2) is *not* sufficient to guarantee that (1.2.3) holds true. Counterexamples, where $G(x)$ has a pathological behaviour around zero, can also be found in [59]. Still, the following holds true in case that ν has a Lebesgue density:

Theorem 1.2.3. *Assume that ν has a Lebesgue density η and that*

$$\liminf_{x\downarrow 0} \frac{\eta(x)+\eta(-x)}{x^{-\alpha-1}} > 0. \tag{1.2.9}$$

Then (1.2.3) holds.

Proof. This result is a direct consequence of Theorem 1.2.1, since (1.2.9) implies (1.2.2). □

A result similar to Theorem 1.2.3 can be shown for Lévy processes which have a moderate activity of small jumps:

Theorem 1.2.4. *Assume that ν has a Lebesgue density η. Assume, moreover, that for some positive constant β and some $\gamma > 0$,*

$$\frac{\eta(x)+\eta(-x)}{|x|^{-1}} \geq \beta - o\left(|x|^{\gamma}\right), x \downarrow 0. \tag{1.2.10}$$

Then we have for some positive constant C_φ:

$$\forall u \in \mathbb{R}: \ |\varphi(u)| \leq C_\varphi (1+|u|)^{-\beta}. \tag{1.2.11}$$

Proof. We can assume, without loss of generality, that $u \geq 0$. Moreover, we can assume that the Gaussian part is zero, since (1.2.11) is automatically satisfied for distributions with nonzero Gaussian component. The absolute value of the characteristic function is

$$|\varphi(u)| = \left|e^{\Psi(u)}\right| = e^{\operatorname{Re}\Psi(u)}. \tag{1.2.12}$$

The real part of the characteristic exponent reads as follows:

$$\operatorname{Re}\Psi(u) = \int (\cos ux - 1)\nu(\,\mathrm{d}x). \tag{1.2.13}$$

Using the fact that $1-\cos ux$ is non-negative and symmetric around zero, we can estimate

$$\begin{aligned}
-\operatorname{Re}\Psi(u) &= \int (1-\cos ux)\nu(\,\mathrm{d}x) \\
&\geq \int_0^1 (1-\cos ux)(\eta(x)+\eta(-x))\,\mathrm{d}x \\
&= \int_0^1 (1-\cos ux)\beta x^{-1}\,\mathrm{d}x - \int_0^1 (1-\cos ux)\left(\beta x^{-1} - (\eta(x)+\eta(-x))\right)\mathrm{d}x \\
&=: \int_0^1 (1-\cos ux)\beta x^{-1}\,\mathrm{d}x - \int_0^1 (1-\cos ux) r(x)\,\mathrm{d}x.
\end{aligned} \tag{1.2.14}$$

Assumption (1.2.10) implies that the following holds true for the positive part of the remainder term:

$$r_+(x) = o(x^{-1+\gamma}),\ x \downarrow 0. \tag{1.2.15}$$

Since, moreover, $r_+(x) \leq \beta x^{-1}$, formula (1.2.15) guarantees integrability of r_+. From this we derive, using the fact that $1 - \cos ux \leq 2$, that for some positive constant C, we have uniformly in u the estimate

$$\begin{aligned} & \int (1-\cos ux)\nu(\,\mathrm{d}x) \\ \geq\ & \beta \int_0^1 (1-\cos ux)x^{-1}\,\mathrm{d}x - 2\int_0^1 r_+(x)\,\mathrm{d}x \\ \geq\ & \beta \int_0^1 (1-\cos ux)x^{-1}\,\mathrm{d}x - C = \beta \int_0^u (1-\cos z)z^{-1}\,\mathrm{d}z - C. \end{aligned} \tag{1.2.16}$$

This implies that for another positive constant C', the following series of inequalities holds uniformly in u:

$$\begin{aligned} & \int (1-\cos ux)\nu(\,\mathrm{d}x) \\ \geq\ & \beta \int_0^u (1-\cos z)z^{-1}\,\mathrm{d}z - C \\ =\ & \beta \int_1^u z^{-1}\,\mathrm{d}z + \int_0^1 (1-\cos z)z^{-1}\,\mathrm{d}z - \int_1^u (\cos z)z^{-1}\,\mathrm{d}z - C \\ \geq\ & \beta \log u - C'. \end{aligned} \tag{1.2.17}$$

We have thus shown that

$$\forall u \in \mathbb{R} :\ -\operatorname{Re}\Psi(u) \geq \beta \log u - C' \tag{1.2.18}$$

and hence, using formula (1.2.13),

$$\forall u \in \mathbb{R} :\ |\varphi(u)| \leq \exp(C')|u|^{-\beta}. \tag{1.2.19}$$

Since we have $\forall u \in \mathbb{R} : |\varphi(u)| \leq 1$, this implies the desired result. □

One can also show that the converse is true: Given a Lévy process of pure jump type, a fast decay of the characteristic function cannot hold unless the jump activity near the origin is high.

Theorem 1.2.5. *Assume that the Gaussian component is zero.*

(i) Assume that for some $\alpha \in (0,2)$, the following holds true:

$$\int_{\{|x|\leq 1\}} |x|^\alpha \nu(\,dx) < \infty. \tag{1.2.20}$$

Then we find that for any positive constant c_φ,

$$\limsup_{|u|\to\infty} \frac{|\varphi(u)|}{\exp\left(-c_\varphi |u|^\alpha\right)} = \infty. \tag{1.2.21}$$

(ii) Assume that

$$\int\limits_{\{|x|\leq\frac{1}{2}\}} \left(\log|x|^{-1}\right)^{-1} \nu(\,dx) < \infty. \tag{1.2.22}$$

Then we find that for arbitrary $\beta > 0$,

$$\limsup_{|u|\to\infty} \frac{|\varphi(u)|}{(1+|u|)^{-\beta}} = \infty. \tag{1.2.23}$$

Proof.

(i) To see the statement of part (i), we show that (1.2.20) implies the following:

$$\liminf_{|u|\to\infty} \frac{-\mathrm{Re}\Psi(u)}{|u|^\alpha} = 0. \tag{1.2.24}$$

Again, it is enough to consider $u \geq 0$. Assume that (1.2.20) holds true. Using Fubini's theorem, we observe that

$$\begin{aligned}\int\limits_{\{|x|\leq 1\}} |x|^\alpha \nu(\,\mathrm{d}x) &= \int\limits_0^1 x^\alpha 2\widetilde{\nu}(\,\mathrm{d}x) = \alpha \int\limits_0^1 \int\limits_0^x z^{\alpha-1}\,\mathrm{d}z 2\widetilde{\nu}(\,\mathrm{d}x) \\ &= \alpha \int\limits_0^1 z^{\alpha-1} \int\limits_z^1 2\widetilde{\nu}(\,\mathrm{d}x)\,\mathrm{d}z = \alpha \int\limits_0^1 z^{\alpha-1} G(z)\,\mathrm{d}z.\end{aligned} \tag{1.2.25}$$

Finiteness of this expression implies

$$\liminf_{z\to 0} \frac{G(z)}{z^{-\alpha}} = 0. \tag{1.2.26}$$

We use the trivial observation that for $z \in (0,1)$ and $\alpha \in (0,2)$, we have $1-\cos z < \frac{z^2}{4} < \frac{z^\alpha}{4}$. Together with (1.2.20), this implies

$$\int\limits_{-\frac{1}{u}}^{\frac{1}{u}} (1-\cos ux)\nu(\,\mathrm{d}x) \leq \frac{1}{4} u^\alpha \int\limits_{-\frac{1}{u}}^{\frac{1}{u}} |x|^\alpha \nu(\,\mathrm{d}x) = u^\alpha o(1),\ u\to\infty. \tag{1.2.27}$$

On the other hand, we have

$$\int\limits_{\frac{1}{u}}^1 (1-\cos ux) 2\widetilde{\nu}(\,\mathrm{d}x) \leq 2 \int\limits_{\frac{1}{u}}^1 2\widetilde{\nu}(\,\mathrm{d}x) = 2G\left(1/u\right). \tag{1.2.28}$$

Together with (1.2.26), this implies that

$$\liminf_{u\to\infty} \frac{\int_{\{|x|\in[\frac{1}{u},1]\}} (1-\cos ux)\nu(\,\mathrm{d}x)}{u^\alpha} = 0. \tag{1.2.29}$$

Finally, we trivially have for the remainder term

$$\int_{\{|x|\geq 1\}} (1-\cos ux)\nu(\,\mathrm{d}x) \leq 2\nu\left((-1,1)^c\right) = O(1),\ u\to\infty. \tag{1.2.30}$$

From (1.2.27), (1.2.29) and (1.2.30) we conclude that

$$\liminf_{|u|\to\infty} \frac{-\mathrm{Re}\Psi(u)}{|u|^\alpha} = 0, \tag{1.2.31}$$

which is the desired result.

(ii) To see the second part of the statement, we can basically argue along the same lines as in the proof of the first part. We replace 1 in the definition of G by $\frac{1}{2}$. Using again Fubini's theorem, we find that

$$\int_{\{|x|\leq\frac{1}{2}\}} \left(\log|x|^{-1}\right)^{-1} \nu(\,\mathrm{d}x) = \int_{\{|x|\leq\frac{1}{2}\}} (\log|x|)^{-2}\,|x|^{-1}G(x)\,\mathrm{d}x. \tag{1.2.32}$$

From (1.2.32) and (1.2.22) we conclude that

$$\liminf_{x\to 0} \frac{G(x)}{\log\left(|x|^{-1}\right)} = 0. \tag{1.2.33}$$

An elementary calculation shows that for any $u \geq e$, we have

$$\forall x\in[-1/2u, 1/2u]:\ (1-\cos ux) \leq \frac{\log u}{\log\left(|x|^{-1}\right)} \tag{1.2.34}$$

(with the convention $\log u/\infty = 0$). From this and from (1.2.22), we conclude that

$$\begin{aligned} &\int_{\{|x|\leq\frac{1}{2u}\}} (1-\cos ux)\nu(\,\mathrm{d}x) \\ \lesssim\ &\log u \int_{\{|x|\leq\frac{1}{2u}\}} \frac{1}{-\log|x|}\nu(\,\mathrm{d}x) = o(1)\log u,\ u\to\infty. \end{aligned} \tag{1.2.35}$$

Moreover, we derive from (1.2.33) that

$$\liminf_{u\to\infty} \frac{\int_{\{|x|\in(1/2u,1/2]\}} (1-\cos ux)\nu(\,\mathrm{d}x)}{\log u} \leq 2\liminf_{u\to\infty} \frac{G(1/2u)}{\log u} = 0. \tag{1.2.36}$$

Again, the remainder term is trivially negligible. We can thus conclude from (1.2.35) and (1.2.36) that

$$\liminf_{|u|\to\infty} \frac{-\Psi(u)}{\log u} = 0 \tag{1.2.37}$$

holds, which implies the desired result. This completes the proof. □

The above results essentially tell us that the smoothness of the Lévy density (if one exists) decreases as the smoothness of the distributional density increases, which may, at the first glance, not be intuitive. Let us have a look at some examples to illustrate this phenomenon.

Examples 1.2.6.

(i) *Let h be the standard normal density. Consider a compound Poisson distribution with intensity λ and standard normally distributed jumps. In this situation, the Lévy density $\eta = \lambda h$ is infinitely differentiable.*

The corresponding infinitely divisible law $\mathbb{P}$ does not even possess a density with respect to the Lebesgue measure, but has point mass at zero.

As the intensity parameter λ gets larger, the mass at the origin vanishes and $\mathbb{P}$ gets "close to having an infinitely differentiable density", see Figure 1.1 below.

(ii) *Consider a gamma distribution with parameters β and λ. The Lévy density*

$$\eta(x) = \beta x^{-1} e^{-\lambda x} 1_{(0,\infty)}(x)$$

has a pole at the origin. The larger β gets , the faster is the decay of η near zero.

The corresponding probability density is

$$g(x) = \frac{\lambda^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\lambda x} 1_{(0,\infty)}.$$

The larger β is, the smoother is g. For $\beta > 1$, $g(x)$ is $\langle\beta\rangle - 1$-times continuously differentiable.

(iii) *Consider an α-stable law: The Lévy density*

$$g(x) = c_1 |x|^{-\alpha-1} 1_{(-\infty,0)}(x) + c_2 |x|^{-\alpha-1} 1_{(0,\infty)}(x)$$

has a pole at the origin. The corresponding probability density is infinitely differentiable.

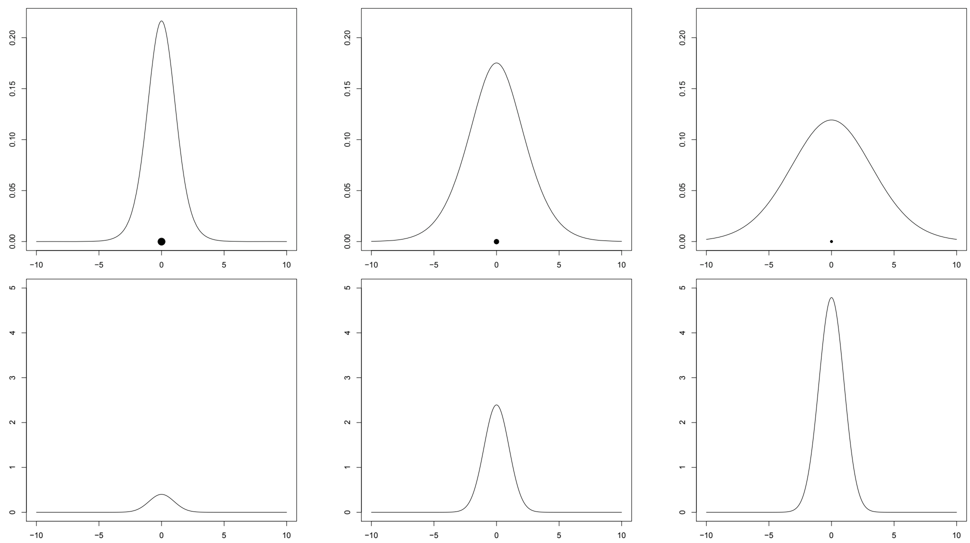

Figure 1.1: Compound Poisson distribution with standard normal distribution of the jumps and intensity parameter $\lambda = 1, 6, 12$. The probability distribution (upper row) gets smoother as the intensity increases. The peak of the Lévy density (second row) gets sharper.

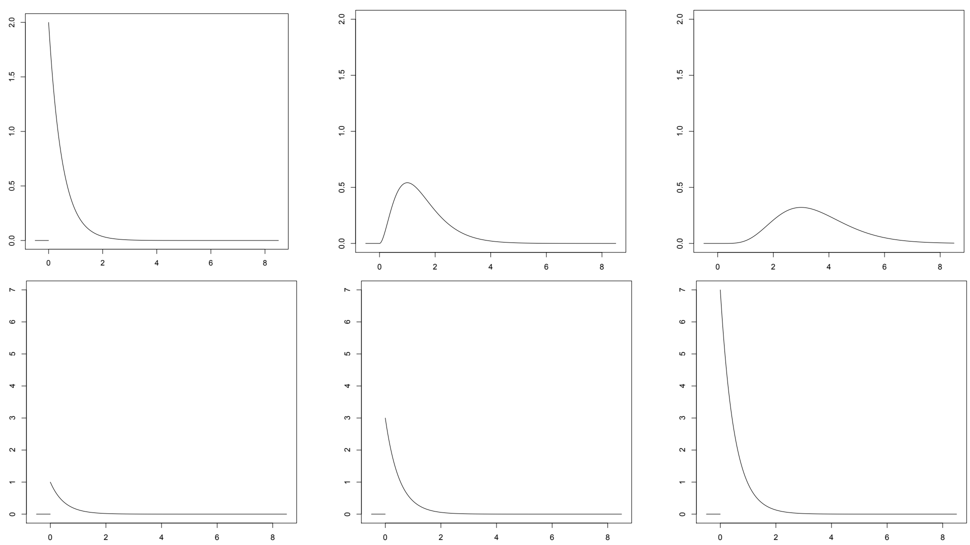

Figure 1.2: Gamma distribution with parameters $\lambda = 2$ and $\beta = 1, 3, 7$. The distributional density (upper row) gets smoother (discontinuous, 2-times continuously differentiable, 6-times continuously differentiable), while the decay of the Lévy density near zero gets faster. The plot in the second row shows $x\eta(x)$.

Chapter 2

Estimation strategy, risk bounds and rates of convergence

For a Lévy processes $X = \{X_t : t \in \mathbb{R}^+\}$ having finite variation on compact sets and finite first moments, $\mu(\mathrm{d}x) := x\nu(\mathrm{d}x)$ is a finite signed measure which completely describes the jump dynamics.

Given low frequency observations of X, we consider the problem of estimating, nonparametrically, some linear functional of μ.

Typical examples are point estimation or estimation of derivatives, if the Lévy measure possesses locally a smooth Lebesgue density. Another example is the estimation of $\nu(A)$ for some compact set $A \subseteq \mathbb{R}$ bounded away from the origin.

Nonparametric estimation of the Lévy measure in presence of low frequency observations of X can be understood as some kind of deconvolution or statistical inverse problem and is intimately connected to nonparametric density deconvolution with unknown error density.

Nonparametric density deconvolution with *known* distribution of the errors is a classical topic in statistics and has been extensively studied in the literature. Optimal rates of convergence for nonparametric deconvolution problems have been studied in the late 80s and early 90s, see Caroll and Hall [16], Stefanski [65], Fan [26] and Efromovich [25], among others. For more recent work on the estimation of linear functionals in the convolution model, we refer to Butucea and Comte [9].

Compared to this, the setting of deconvolution with *unknown* distribution of the errors is still of independent interest and a topic of ongoing research. We refer here, at the first place, to the work by Neumann, starting in the late 90s, see [57] and [55], to Johannes [36] and to Comte and Lacour [20].

Motivated by this fact, we shall use the classical setting of density deconvolution as a toy model to better understand the problem at hand and develop our ideas not only for the particular problem of nonparametric estimation for Lévy processes, but also sketch the application to density deconvolution with unknown error distribution.

The parameter of interest is estimated by Fourier methods, replacing the unknown characteristic function and its derivative by their empirical counterparts.

Estimation via Fourier methods is standard in nonparametric density deconvolution. The application to low frequently observed Lévy processes has been investigated in 2009 by Neumann and Reiß [56]. In that paper, the estimator is obtained as the solution of an abstract minimisation problem over classes of finite measures.

Compared to this, the estimation procedure presented in the present section is more in the spirit of the constructive estimators proposed by Comte and Genon-Catalot in [18] and by Gugushvili [33, 34]. However, we find that the spectral-cutoff estimator proposed in those papers suffers from serious drawbacks when applied to the problem of estimating linear functionals of the Lévy density. This is due to the structural properties of a Lévy measure, compared to a usual probability measure. For this reason, we will have to localise the procedure and work with general kernel functions.

The present chapter is organised as follows: In Section 2.1, we describe the statistical model and technical assumptions. In Section 2.2, we start by considering the problem of estimating a linear functional in the density deconvolution model with unknown error distribution. Inspired by the reasoning given by Butucea and Comte in [9], we introduce spectral cutoff estimators and derive upper bounds on the corresponding risk. Next, we discuss the problems which arise when trying to extend this reasoning to the Lévy model and introduce general kernel estimators. In Section 2.3, we derive rates of convergence under regularity assumptions on the Lévy measure ν and on the decay of the characteristic function. Finally, in Section 2.4 we derive lower bound for point estimation and estimation of integrals and see that the rates of convergence which are found in Section 2.3 are indeed minimax optimal.

Most of the proofs are postponed to Section 2.5.

2.1 Statistical model and assumptions

We consider the following statistical model:

Model 1 (Lévy model)**.** *A Lévy process $X = \{X_t : t \in \mathbb{R}^+\}$ is observed at equidistant time points $\Delta, \cdots, 2n\Delta =: \Delta, \cdots, 2T$. This is equivalent to stating that we observe $2n$ i.i.d. copies*

$$Z_{\Delta,j} := X_{j\Delta} - X_{(j-1)\Delta}, \; j = 1, \cdots, 2n \tag{2.1.1}$$

of the infinitely divisible random variable X_Δ.

Throughout this chapter, we assume that we are in the low frequency regime, that is, the distance Δ between the observation times is fixed.

Our goal is to infer the underlying jump dynamics.

When dealing with continuous time observations, the way to go is to use the empirical jump measure as an estimator of the true underlying jump measure. The expected number of jumps per time unit in some measurable set bounded away from zero is replaced by the observed number of jumps. When estimating the corresponding Lebesgue density, one has to apply some smoothing procedure, typically projection estimators. This approach has been investigated in [30].

Next, when placing oneself in a high frequency model, that is, when assuming that the distance Δ between the observation times tends to zero at a high enough rate, one might discretise the procedure. The jumps are no longer directly feasible, but are observable in the limit. Roughly speaking, a large increment $X_{s+\Delta} - X_s$ within a small time interval will be due to a large jump. When observing the process at a high enough frequency, one is eventually able to "see" the jumps. For the details, we refer to [29] and subsequent papers and to [17, 19] and the discussion therein.

Given low frequency observations, the jumps are latent, unobservable quantities and one deals with a more complicated deconvolution type problem. In this setting, we will have to exploit the structure of infinitely divisible distributions, namely the connection between the underlying Lévy measure and the characteristic function. This approach has first been investigated in [56] and then in [18] and in [33] and [34].

We work under the following technical assumptions on the process X under consideration:

Assumptions 2.1.1.

(A1) *X is of pure jump type.*

(A2) *X has moderate jump activity in the sense that the following holds for the Lévy measure:*

$$\int_{-1}^{1} |x|\nu(\,dx) < \infty. \tag{2.1.2}$$

(A3) *X does not have a drift component.*

(A4) *For one and hence for any $t > 0$, X_t has a finite second moment. This is equivalent to stating that*

$$\int x^2\nu(\,dx) < \infty. \tag{2.1.3}$$

Imposing the assumptions *(A1)* and *(A2)* is equivalent to stating that X has finite variation on compact sets.

It was seen in Section 1.1 that prototypical examples are compound Poisson processes, gamma processes and tempered stable processes without drift component and with index $\alpha < 1$.

It is worth mentioning that most of these conditions can be relaxed. It would be possible to allow a Gaussian component and a high activity of small jumps. However, these modifications will result in a more complicated structure of the estimator and the proofs without allowing much further insight into the problem, for which reason we work under the above assumptions. We give a brief discussion about a more general model in Appendix C.

Let φ_Δ denote the characteristic function of X_Δ. It was shown in Section 1.1 that under the above assumptions, we have the following particular form of the Lévy-Khintchine representation: The characteristic function is given by $\varphi_\Delta = \exp(\Delta\Psi)$, where the characteristic exponent reads as follows:

$$\Psi(u) = \int (e^{iux} - 1)\nu(\,\mathrm{d}x) = \int \frac{e^{iux} - 1}{x} x\nu(\,\mathrm{d}x). \tag{2.1.4}$$

The process is thus fully described by the signed measure $\mu(\mathrm{d}x) := x\nu(\mathrm{d}x)$, which is finite thanks to (2.1.2) and (2.1.3).

Our goal is to estimate some linear functional of μ. Let some distribution f be given. Then the quantity of interest is

$$\theta := \langle f, \mu \rangle := \int f(x)\mu(\mathrm{d}x). \tag{2.1.5}$$

(The notation which we use here is, in some cases, not canonical and the integral appearing in formula (2.1.5) may be confusing. We refer, at this point, to Appendix B for explanation.)

We assume henceforth that f satisfies one of the following assumptions:

Assumptions 2.1.2.

(F1) f is a regular distribution, $f \in \mathrm{L}^1(\mathbb{R})$ and $\|f\|_\infty := \sup_{x\in\mathbb{R}} |f(x)| < \infty$.

(F2) f is compactly supported.

In case that *(F1)* is true, boundedness of f and finiteness of μ clearly imply that the integral appearing in (2.1.5) converges so the definition of θ makes sense. For non-regular f, we formulate an additional assumption on μ which makes the problem well defined.

Assumption 2.1.3. *If f is non-regular and compactly supported and k denotes the order of f (see Definition B.5), the following holds true for μ:*

(A5) For some open interval $D = (d_1, d_2)$ with $\operatorname{supp}(f) \subseteq D$, the restriction $\mu|_D$ possesses a density $g_D \in C^k(D)$.

Consider the following typical problems:

Examples 2.1.4.

(i) Point estimation: If the Lévy measure ν possesses (locally) a continuous Lebesgue density $\eta(x)$, one is often interested in the value of η at some point $y \in \mathbb{R}\backslash\{0\}$. Let $f := \delta_y$ denote the Dirac distribution. Then the parameter of interest is

$$\theta = \eta(y) = y^{-1} \int \delta_y(x)\mu(\,dx).$$

Moreover, in case that the jump density is differentiable at some point, one might be interested in the value of the derivative.

(ii) Given some compact set $A \subseteq \mathbb{R}\backslash\{0\}$, the expected number of jumps per time unit with size in A might be of interest. In this case, we consider the test function $f(x) = x^{-1}1_A(x)$, so

$$\theta = \nu(A) = \int x^{-1}1_A(x)\mu(\,dx).$$

(iii) *In applications, one is often interested in estimating the moments of μ, that is, in*

$$\theta = \int x^m \mu(\,dx) =: \int f(x)\mu(\,dx).$$

In this case, f is certainly not compactly supported nor bounded or integrable. Still, one might consider a truncated or tempered version of f.

Formula (2.1.4) allows to recover the Fourier transform $\mathcal{F}\mu$ of μ by derivating the characteristic exponent. Dominated convergence permits to derivate under the integral sign, so we have

$$\begin{aligned}\Psi'(u) = \quad & \frac{\partial}{\partial u}\int \frac{(e^{iux}-1)}{x}\mu(\,\mathrm{d}x) \\ = \quad & \int \frac{\partial}{\partial u}\frac{(e^{iux}-1)}{x}\mu(\,\mathrm{d}x) = i\int e^{iux}\mu(\,\mathrm{d}x) = i\mathcal{F}\mu(u).\end{aligned} \tag{2.1.6}$$

In terms of the characteristic function and its derivative this can be expressed as follows:

$$\mathcal{F}\mu(u) = \frac{1}{\Delta}(-i\Delta\Psi'(u)) = \frac{\frac{1}{\Delta}\varphi_\Delta'(u)}{i\varphi_\Delta(u)}. \tag{2.1.7}$$

Recall that the characteristic function of an infinitely divisible law possesses no zeros, so dividing by φ_Δ is not critical.

Formula (2.1.7) indicates a strong structural resemblance to a nonparametric density deconvolution problem. Let us recall the setting:

Model 2 (Density deconvolution model)**.** *We observe*

$$Z_j = Y_j + \varepsilon_j,\ j = 1, \cdots, n, \tag{2.1.8}$$

where (Y_j) and (ε_j) are independent sequences of real valued random variables, the Y_j are are i.i.d. with unknown distribution $\mathbb{P}^Y$ and the ε_j are i.i.d. with density f_ε, which is unknown.

In this model, the distribution $\mathbb{P}^Z$ of the Z_j is the convolution of $\mathbb{P}^Y$ and f_ε. In case that $\mathcal{F}f_\varepsilon$ has no zeros, this allows to express the Fourier transform of $\mathbb{P}^Y$ as follows:

$$\mathcal{F}\mathbb{P}^Y(u) = \frac{\mathcal{F}\mathbb{P}^Z(u)}{\mathcal{F}f_\varepsilon(u)}. \tag{2.1.9}$$

The problem of estimating $\mathbb{P}^Y$ in presence of noisy observations of the Y_j is a prototypical statistical inverse problem: The degree of ill-posedness will depend on the decay of $\mathcal{F}f_\varepsilon$ as well as on the decay of $\mathcal{F}\mathbb{P}^Y$. The faster $\mathcal{F}f_\varepsilon$ decays, the more complicated will it be to recover $\mathbb{P}^Y$ from the data, since for small values of $\mathcal{F}f_\varepsilon$, even a drastic change in $\mathcal{F}\mathbb{P}^Y$ may not lead to a large change of $\mathcal{F}\mathbb{P}^Z$. On the other hand, the faster $\mathcal{F}\mathbb{P}^Y$ decays the better rates of convergence can be obtained. (For a more general discussion on rates of convergence for statistical inverse problems, see, for example, [31].)

On an intuitive level, the dependence between the classical model of nonparametric density deconvolution and the problem of nonparametric estimation for

Lévy processes can be understood as follows: Within a small time interval Δ, an increment $X_{s+\Delta} - X_s$ with magnitude in some compact set $A \subseteq \mathbb{R}\setminus\{0\}$ will mostly be due to some big jump with size in A, perturbed by a (possibly) infinite number of small jumps. The higher the jump activity is, the stronger is the smoothing effect due to the small jumps and the more complicated it is to recover the jump density. The jump measure away from zero thus plays, roughly speaking, the role of the unknown probability distribution $\mathbb{P}^Y$ in the density deconvolution model, while the degree of ill-posedness is ruled by the degree of activity of small jumps, which was shown to determine the decay of the characteristic function (see Section 1.2). For large Δ and arbitrary $n \in \mathbb{N}$, we have $\mathbb{P}_\Delta = \left(\mathbb{P}_{\Delta/n}\right)^{*n}$, so the smoothness properties for a small time horizon determine also the smoothness of $\mathbb{P}_\Delta$ for large values of Δ.

In formula (2.1.7), the derivative of the characteristic function appearing in the numerator is simply the Fourier transform of the finite signed measure $x\,\mathbb{P}_\Delta(\,\mathrm{d}x) := x\,\mathbb{P}^{X_\Delta}(\,\mathrm{d}x)$, multiplied by i, which is easily seen by noting that (using again dominated convergence)

$$\begin{aligned}\varphi'_\Delta(u) &= \frac{\partial}{\partial u}\mathbb{E}\left[e^{iuX_\Delta}\right] = \mathbb{E}\left[\frac{\partial}{\partial u}e^{iuX_\Delta}\right] = \mathbb{E}\left[iX_\Delta e^{iuX_\Delta}\right] \\ &= i\int e^{iux}x\,\mathbb{P}_\Delta(\,\mathrm{d}x) = i\Big(\mathcal{F}\left(x\,\mathbb{P}_\Delta(\,\mathrm{d}x)\right)\Big)(u). \end{aligned}\tag{2.1.10}$$

We do thus have a convolution structure in the sense that our quantity of interest μ is given by the convolution equation

$$x\,\mathbb{P}_\Delta(\,\mathrm{d}x) = (\Delta\mu)(\,\mathrm{d}x) * \mathbb{P}_\Delta(\,\mathrm{d}x). \tag{2.1.11}$$

Asymptotically, with delta tending to zero, the convolution structure is lost. $\mathbb{P}_\Delta$ degenerates to the Dirac measure at zero and $\frac{1}{\Delta}x\,\mathbb{P}_\Delta$ can be shown to tend weakly to μ. This explains why one recovers, in the high frequency regime, the rates of convergence which are classical in density estimation (see [17, 19] and [29]). For the details on small time asymptotic properties of Lévy processes, we refer to [28, 27] and to [19].

However, we assume, in the present setting, that the process is observed at low frequency, that is, Δ is fixed.

The model (2.1.8) of nonparametric deconvolution is quite well studied in the literature. Point estimation is a very classical topic and the problem of estimating general linear functionals in the convolution model is considered in [9]. However, it is assumed there that the distribution of the errors is known.

Contrarily to this, the characteristic function φ_Δ appearing in the denominator in formula (2.1.7) is clearly unknown and the same is assumed to be true in Model 2.

2.2 Estimation procedure and risk bounds

Estimation in the convolution model

First, we place ourselves in the Model 2 of density deconvolution with unknown distribution of the errors and consider the problem of estimating a linear functional $\theta = \langle f, \mathbb{P}^Y \rangle$ of the underlying distribution $\mathbb{P}^Y$. We assume, in the sequel, that one of the assumptions *(F1)* or *(F2)* is satisfied.

We impose, moreover, the following assumptions:

Assumptions 2.2.1.

(D1) There are n i.i.d. observations $\varepsilon_{-n}, \cdots, \varepsilon_{-1} \sim f_\varepsilon$ of the pure noise available.

(D2) The following holds true for the Fourier transform $\mathcal{F}f$ of the distribution f and the Fourier transform $\mathcal{F}\mathbb{P}^Y$ of $\mathbb{P}^Y$:

$$\mathcal{F}f \mathcal{F}\mathbb{P}^Y \in \mathrm{L}^1(\mathbb{R}). \tag{2.2.1}$$

(D3) The characteristic function $\mathcal{F}f_\varepsilon$ of the noise has no zeros.

(D4) If (F2) is met and k denotes the order of f, $\mathbb{P}^Y$ has a density $g_D \in C^k(D)$ in a neighbourhood D of $\operatorname{supp}(f)$.

A remark is in order, concerning assumption *(D2)*. Since f is a distribution, that is, a "generalised function", there may be confusion about what kind of object $\mathcal{F}f$ is. If *(F1)* is true, it is clear that $\mathcal{F}f$ is the usual Fourier transform. If *(F2)* is met, Theorem B.14 in Appendix B will tell us that $\mathcal{F}f$ is indeed a regular distribution and can be identified with a smooth function. For this reason, formula (2.2.1) makes sense.

Under the above assumptions, θ is well defined and can be expressed in the Fourier domain, using Parseval's identity (see Theorem B.13 in the appendix):

$$\theta = \int f(x)\, \mathbb{P}^Y(\,\mathrm{d}x) = \frac{1}{2\pi} \int \mathcal{F}f(-u) \mathcal{F}\mathbb{P}^Y(u)\, \mathrm{d}u. \tag{2.2.2}$$

Formula (2.1.9) gives

$$\frac{1}{2\pi} \int \mathcal{F}f(-u) \mathcal{F}g_Y(u)\, \mathrm{d}u = \frac{1}{2\pi} \int \mathcal{F}f(-u) \frac{\mathcal{F}\mathbb{P}^Z(u)}{\mathcal{F}f_\varepsilon(u)}\, \mathrm{d}u. \tag{2.2.3}$$

This suggests to estimate θ by Fourier methods, replacing $\mathcal{F}\mathbb{P}^Z =: \varphi_Z$ and $\mathcal{F}f_\varepsilon =: \varphi_\varepsilon$ by their empirical counterparts. This is done by setting

$$\widehat{\varphi}_{Z_n}(u) := \frac{1}{n} \sum_{j=1}^{n} e^{iuZ_j} \tag{2.2.4}$$

and

$$\widehat{\varphi}_{\varepsilon_n}(u) := \frac{1}{n} \sum_{j=-n}^{-1} e^{iu\varepsilon_j}. \tag{2.2.5}$$

Next, $\widehat{\varphi}_{\varepsilon_n}$ is replaced in the denominator by its truncated version: Using the definition which is originally due to Neumann (see [57]), we set

$$\frac{1}{\widetilde{\varphi}_{\varepsilon_n}(u)} := \frac{1\left(\{|\widehat{\varphi}_{\varepsilon_n}(u)| \geq n^{-1/2}\}\right)}{\widehat{\varphi}_{\varepsilon_n}(u)}. \tag{2.2.6}$$

The following key result which has been proved in [57] for $k = 2$ and immediately generalises to other values of k gives control on the deviation of $\frac{1}{\widetilde{\varphi}_{\varepsilon_n}(u)}$ from its target:

Lemma 2.2.2. *Let $\frac{1}{\widetilde{\varphi}_{\varepsilon_n}}$ be defined by (2.2.6). Then we have for arbitrary $k \in \mathbb{N}$ and universal constants C_k depending only on k:*

$$\mathbb{E}\left[\left|\frac{1}{\widetilde{\varphi}_{\varepsilon_n}(u)} - \frac{1}{\varphi_\varepsilon(u)}\right|^k\right] \leq C_k\left(\frac{n^{-k/2}}{|\varphi_\varepsilon(u)|^{2k}} \wedge \frac{1}{|\varphi_\varepsilon(u)|^k}\right) \tag{2.2.7}$$

If f has integrable Fourier transform, we can define a direct plug-in estimator:

Definition 2.2.3. *In Model 2, assume that Assumptions 2.2.1 are met. Assume, moreover, that $\mathcal{F}f \in \mathrm{L}^1(\mathbb{R})$. Then we set*

$$\widehat{\theta}_n := \frac{1}{2\pi}\int \mathcal{F}f(-u)\frac{\widehat{\varphi}_{Z_n}(u)}{\widetilde{\varphi}_{\varepsilon_n}(u)}\, du. \tag{2.2.8}$$

This definition is meaningful since the absolute values of $\widehat{\varphi}_{Z_n}$ and $\frac{1}{\widetilde{\varphi}_{\varepsilon_n}}$ are by definition bounded above and $\mathcal{F}f$ is integrable by assumption, so the integral appearing in (2.2.8) is well defined and finite.

However, in many cases, the Fourier transform of f is not integrable so the integral appearing in (2.2.8) generally fails to converge. This is true, for example, when one considers point estimation, that is, when $f = \delta_y$.

In this case, one will have to apply some additional smoothing procedure. For this purpose, it is often convenient to work with a spectral cutoff estimator. The following definition is in accordance with the reasoning presented in [9].

Definition 2.2.4. *In Model 2, assume that Assumptions 2.2.1 are met. Assume, moreover, that (F1) or (F2) is satisfied. We define for $m \in \mathbb{N}$:*

$$\widehat{\theta}_{m,n} := \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \mathcal{F}f(-u)\frac{\widehat{\varphi}_{Z_n}(u)}{\widetilde{\varphi}_{\varepsilon_n}(u)}\, du. \tag{2.2.9}$$

Since $\mathcal{F}f$ is locally bounded and $\left|\frac{\widehat{\varphi}_{Z_n}}{\widetilde{\varphi}_{\varepsilon_n}}\right|$ is bounded above by definition, the integral appearing in formula (2.2.9) is always well defined an finite.

We can give the following bounds on the risk of the estimators thus defined.

Theorem 2.2.5. *Let i.i.d. observations $Z_1, \cdots, Z_n$ according to Model 2 be given. Assume that (D1)-(D4) and (F1) or (F2) are satisfied. Let $\widehat{\theta}_{m,n}$ be the* *estimator introduced in Definition 2.2.4. Then we can estimate*

$$
\begin{aligned}
&\mathbb{E}\left[\left|\theta - \widehat{\theta}_{m,n}\right|^2\right] \\
&\leq \frac{1}{2\pi^2}\left| \int_{\{|u|\geq \pi m\}} \mathcal{F}f(-u)\varphi_Y(u)(u)\, du \right|^2 \\
&+ \frac{n^{-1}}{2\pi^2}\left\{ C_D \int_{\{|u|\leq \pi m\}} \left|\frac{\mathcal{F}f(-u)}{\varphi_\varepsilon(u)}\right|^2 du \wedge 3C \left(\int_{\{|u|\leq \pi m\}} \left|\frac{\mathcal{F}f(-u)}{\varphi_\varepsilon(u)}\right| du \right)^2 \right\}, \quad (2.2.10)
\end{aligned}
$$

with constant

$$
C_D = C\left(\int |\varphi_Y(x)|\, dx + 2\int |\varphi_Y(x)|^2\, dx \right) \leq \infty. \quad (2.2.11)
$$

C is some positive constant depending only on the constants C_1 and C_2 in Lemma 2.2.2.

The occurrence of the constant C_D is in accordance with the structure found in [9], but the value is different. This results from estimating the unknown characteristic function in the denominator.

In case that f has integrable Fourier transform we obtain the following upper bound on the risk of the estimator $\widehat{\theta}_n$:

Theorem 2.2.6. *In the situation of the preceding theorem, assume that $\mathcal{F}f \in \mathrm{L}^1(\mathbb{R})$. Then the following estimate holds true for $\widehat{\theta}_n$ defined as in Definition 2.2.4: For arbitrary $m \geq 0$, we have*

$$
\begin{aligned}
&\mathbb{E}\left[\left|\theta - \widehat{\theta}_{n}\right|^2\right] \\
&\leq \frac{1}{2\pi^2}\left\{ C_D \int_{\{|u|>\pi m\}} |\mathcal{F}f(-u)|^2\, du \wedge 3C \left(\int_{\{|u|>\pi m\}} |\mathcal{F}f(-u)|\, du \right)^2 \right\} \\
&+ \frac{n^{-1}}{2\pi^2}\left\{ C_D \int_{\{|u|\leq \pi m\}} \left|\frac{\mathcal{F}f(-u)}{\varphi_\varepsilon(u)}\right|^2 du \wedge 3C \left(\int_{\{|u|\leq \pi m\}} \left|\frac{\mathcal{F}f(-u)}{\varphi_\varepsilon(u)}\right| du \right)^2 \right\}, \quad (2.2.12)
\end{aligned}
$$

with C_D defined as in Theorem 2.2.5.

We face, in Theorem 2.2.5 the usual trade off between the approximation error which results from cutting off in the Fourier domain and the error within the model. Choosing m large reduces the approximation error, but at the cost of making the error in the model explode. On the other hand, choosing m small will reduce the error in the model, but at the cost of producing a large bias.

The result given in Theorem 2.2.5 is easily seen to carry over to the simpler setting of density deconvolution with *known* distribution of the noise. In this setting, one can (artificially) truncate the known characteristic function φ_ε at the threshold value $n^{-1/2}$, which will lead to a well defined estimator if $\mathcal{F}f$ is integrable. However, Theorem 2.2.5 shows that better rates of convergence can be derived for the spectral cutoff estimator if $\mathbb{P}^Y$ possesses a smooth Lebesgue density.

Estimation in the Lévy model

It seems natural to apply the above procedure to our original problem of estimating the jump measure, when given low frequency observations of a pure jump Lévy process. Spectral cutoff estimators for Lévy processes observed at high frequency (with L^2-loss) have been considered, for example, in [18] and in [33].

By the definition of a Lévy process, the increments form i.i.d. copies of X_Δ. We can thus define the empirical versions of φ_Δ and φ'_Δ, setting

$$\widehat{\varphi}_{\Delta,n}(u) := \frac{1}{n}\sum_{k=1}^{n} e^{iuZ_{\Delta,k}} \tag{2.2.13}$$

and

$$\widehat{\varphi}'_{\Delta,n}(u) := \frac{1}{n}\sum_{k=n+1}^{2n} iZ_{\Delta,k}e^{iuZ_{\Delta,k}}. \tag{2.2.14}$$

Again, truncating the empirical characteristic function in the denominator, we set

$$\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)} := \frac{1(|\widehat{\varphi}_{\Delta,n}(u)| \geq (\Delta n)^{-1/2})}{\widehat{\varphi}_{\Delta,n}(u)}. \tag{2.2.15}$$

Given a distribution f, we might follow the definition in formula (2.2.9) and use

$$\widehat{\theta}_{\Delta,m,n} := \int\limits_{\{|u|\leq \pi m\}} \mathcal{F}f(-u)\frac{\frac{1}{\Delta}\widehat{\varphi}'_{\Delta,n}(u)}{i\widetilde{\varphi}_{\Delta,n}(u)}\,\mathrm{d}u \tag{2.2.16}$$

as an estimator of $\theta = \langle f, \mu\rangle$.

When working under the condition *(F1)* that f is a bounded and integrable function, the condition $\mathcal{F}f\mathcal{F}\mu \in \mathrm{L}^1(\mathbb{R})$, which is related to *(D2)*, holds under mild regularity assumptions on μ, for example, under the assumption that μ has a square integrable Lebesgue density.

In this case, the identity

$$\theta = \frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\mu(u)\,\mathrm{d}u \tag{2.2.17}$$

holds true and working with the spectral cutoff estimator makes sense.

However, when considering point estimation or estimation of derivatives at some point $y \in \mathbb{R}\setminus\{0\}$ or, more abstract, when assuming that f is a non-regular distribution, this approach is no longer satisfactory.

To see this, we have to take into account the structural properties of a Lévy density compared to a usual probability density. When considering the rates of convergence which can be derived from Theorem 2.2.5 as functions of the decay of φ_ε, and the smoothness of f and of the density g_Y of $\mathbb{P}^Y$, it is clear that the appropriate concept for measuring the smoothness is (global) Sobolev regularity. The faster the decay of $\mathcal{F}g_Y$ the easier it is to recover the density g_Y.

However, when measured in a Sobolev sense, μ cannot have a very smooth Lebesgue density g, except from the particular case of compound Poisson processes. Whenever X has infinite jump activity, g will have a point of discontinuity at zero.

Consider, for example, Gamma processes. The Lévy measure ν has a density η which behaves as $|x|^{-1}$ at zero, so the density g of μ has a jump at the origin. For the Fourier transform $\mathcal{F}\mu$ of μ, we have $|\mathcal{F}\mu(u)| \sim |u|^{-1},\ |u| \to \infty$.

For processes having stable like behaviour at the origin, we have $\eta(x) \sim |x|^{-1-\alpha},\ x \to 0$, so g is unbounded and behaves as $|x|^{-\alpha}$ at the origin. The Fourier transform $\mathcal{F}\mu$ decays as $|u|^{-1+\alpha}$ as $|u|$ tends to infinity.

Consequently, for Lévy processes having infinite jump activity, on can usually not assume that $\mathcal{F}f\mathcal{F}\mu$ is integrable and the identity $\int f(x)\mu(\mathrm{d}x) = \frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\mu(\mathrm{d}x)$ generally fails to hold. The spectral cutoff estimator given in (2.2.16) is still well defined and finite, but need not even be consistent as an estimator of θ.

In density deconvolution, the assumption that the underlying Lebesgue density g_Y is globally smooth, say, continuously differentiable, is not a big restriction, whereas μ is globally smooth only in exceptional cases.

Indeed, when dealing with point estimation or, more generally, with the estimation of linear functionals defined on compact sets bounded away from the origin, the appropriate concept to take into account is *local* (Hölder) regularity in some neighbourhood of the point or interval of interest. This means that one will have to localise the procedure and work with general kernel functions which decay at a high enough rate, thus excluding the influence of the singularity at zero. For this purpose, the sinc kernel is usually inappropriate.

This leads to considering general kernel functions rather than limiting the considerations to the spectral cutoff estimator.

In the sequel, we let K be some kernel function on which we impose the following conditions:

(K1) For any $h > 0$, we have $\mathcal{F}\mathrm{K}_h\,\mathcal{F}f(-\bullet) \in \mathrm{L}^1(\mathbb{R})$.

(K2) If f is non-regular with order k, K is $k-$ times continuously differentiable.

Replacing the sinc kernel in (2.2.16) by a general kernel function leads to the following definition:

Definition 2.2.7. *In Model 1, let Assumptions 2.1.1 be satisfied. Assume that (F1) or (F2) is met and, in addition, Assumption 2.1.3 is satisfied. Let K be chosen such that (K1) and (K2) are met.*

For a bandwidth $h > 0$, we define

$$\widehat{\theta}_{\Delta,h,n} := \frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\mathrm{K}_h(u)\frac{\frac{1}{\Delta}\widehat{\varphi}'_{\Delta,n}(u)}{i\widetilde{\varphi}_{\Delta,n}(u)}\,du. \tag{2.2.18}$$

Since $\left|\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}\right|$ is bounded above by definition, assumption *(K1)* guarantees that $\widehat{\theta}_{\Delta,h,n}$ is well defined and finite.

If f has an integrable Fourier transform, we can in analogy with Definition 2.2.3, define an estimator without additional smoothing.

Definition 2.2.8. *In the situation of Definition 2.2.7, assume that $\mathcal{F}f \in \mathrm{L}^1(\mathbb{R})$. Then we set*

$$\widehat{\theta}_{\Delta,n} := \frac{1}{2\pi} \int \mathcal{F}f(-u) \frac{\frac{1}{\Delta}\widehat{\varphi}'_{\Delta,n}(u)}{i\widetilde{\varphi}_{\Delta,n}(u)}\, du. \tag{2.2.19}$$

We can give the following risk bounds:

Theorem 2.2.9. *Let observations $X_\Delta, \cdots, X_{2\Delta}$ according to Model 1 be given. Assume that (A1)- (A4) are satisfied. Let f be a distribution for which (F1) or (F2) is met. Let K be some kernel function for which (K1) and (K2) hold true. For non-regular f, assume, moreover, that (A5) is satisfied. For $h > 0$, let $\widehat{\theta}_{\Delta,h,n}$ be defined by (2.2.18). Then we can estimate*

$$\begin{aligned}
&\mathbb{E}\left[\left|\theta - \widehat{\theta}_{\Delta,h,n}\right|^2\right] \\
&\leq 2\left|\int f(x)\mu(\,dx) - \int f(x)\,(\mathrm{K}_h * \mu)\,(\,dx)\right|^2 \\
&+ \frac{T^{-1}}{2\pi^2}\left\{C_1\int |\mathcal{F}\mathrm{K}(hu)|^2 \left|\frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)}\right|^2 du \wedge C_2\left(\int |\mathcal{F}\mathrm{K}(hu)|\left|\frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)}\right| du\right)^2\right\},
\end{aligned} \tag{2.2.20}$$

with constants

$$C_1 = C\left(\int |\Psi''(x)|\, dx + 2\int |\Psi'(x)|^2\, dx\right) \leq \infty \tag{2.2.21}$$

and

$$C_2 = C\left(\|\Psi''\|_\infty + 2\|\Psi'\|_\infty^2\right) < \infty. \tag{2.2.22}$$

C is a positive constant depending only on the constants C_1 and C_2 in Lemma 2.2.2.

For the estimator which is obtained without smoothing, we can give the following risk bound:

Theorem 2.2.10. *In the situation of the preceding theorem, assume that $\mathcal{F}f \in \mathrm{L}^1(\mathbb{R})$. Let the estimator $\widehat{\theta}_{\Delta,n}$ of θ be defined according to Definition 2.2.8. Then we have the following bound on the squared risk of $\widehat{\theta}_{\Delta,n}$:*

For arbitrary $m \geq 0$, we can estimate

$$\begin{aligned}
&\mathbb{E}\left[\left|\theta - \widehat{\theta}_{\Delta,n}\right|^2\right] \\
\leq\ & \frac{1}{2\pi^2}\left\{C_1 \int\limits_{\{|u|>\pi m\}} |\mathcal{F}f(-u)|^2\, du \wedge C_2 \left(\int\limits_{\{|u|>\pi m\}} |\mathcal{F}f(-u)|\, du\right)^2\right\} \\
+\ & \frac{T^{-1}}{2\pi^2}\left\{C_1 \int\limits_{\{|u|\leq\pi m\}} \frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2}\, du \wedge C_2 \left(\int\limits_{\{|u|\leq\pi m\}} \frac{|\mathcal{F}f(-u)|}{|\varphi_\Delta(u)|}\, du\right)^2\right\}, \quad (2.2.23)
\end{aligned}$$

with constants C_1 and C_2 defined as in Theorem 2.2.9.

It should be noted that the statement of Theorem 2.2.10 is in analogy with the main result obtained in [56]. Indeed, the minimum-distance estimator which has been introduced by Neumann and Reiß corresponds to the estimator which is given by Definition 2.2.8.

The obvious advantage of our approach is that $\widehat{\theta}_{\Delta,n}$ is a constructive estimator and can be calculated directly from the data. We can thus circumvent the abstract minimisation problem over classes of measures, which arises in [56]. This is certainly comfortable in applications.

Moreover, the minimum distance fit is a priori not appropriate for pointwise estimation or estimating derivatives and it is not obvious how a kernel smoothing procedure can be included in this setting.

2.3 Rates of convergence

In the present section, we investigate the rates of convergence which can be derived from the upper risk bounds given in Theorem 2.2.9 and Theorem 2.2.10 under regularity assumptions on μ as well as on f and on $\mathbb{P}_\Delta$.

Let us first introduce the following abstract nonparametric classes:

Definition 2.3.1.

(i) We denote by $\mathcal{F}(\beta, \rho, D_1, D_2, d_1, d_2)$ the class of functions f such that for any $u \in \mathbb{R}$:

$$D_1(1+|u|)^{-\beta} \exp\left(-d_1|u|^\rho\right) \leq |\mathcal{F}f(u)| \leq D_2(1+|u|)^{-\beta} \exp\left(-d_2|u|^\rho\right). \quad (2.3.1)$$

If $\rho = 0$ and $\beta > 0$, the functions collected in $\mathcal{F}(\beta, \rho, D_1, D_2, d_1, d_2)$ are called ordinary smooth. *For $\rho > 0$, they are called* supersmooth.

(ii) Given $a > 0$, let $\langle a \rangle := \sup\{k \in \mathbb{N} : k < a\}$. For an open subset $D \subseteq \mathbb{R}$, we denote by $\mathcal{H}_D(\alpha, L, R)$ the class of functions f such that $\sup_{x \in D} |f(x)| \leq R$, $f|_D$ is $\langle a \rangle$ times continuously differentiable and we have

$$\sup_{\substack{x,y \in D \\ x \neq y}} |f^{(\langle a \rangle)}(x) - f^{(\langle a \rangle)}(y)| \leq L|x-y|^{a - \langle a \rangle}. \quad (2.3.2)$$

The functions belonging to $\mathcal{H}_D(a, L, R)$ *are called locally* Hölder regular *with index* a*.*

(iii) *For* $a, M \geq 0$*, the* Sobolev class $\mathcal{S}(a, M)$ *consists of all functions* $f \in \mathrm{L}^2(\mathbb{R})$ *such that*

$$\int (1+|u|^2)^a |\mathcal{F}f(-u)|^2\, du \leq M. \tag{2.3.3}$$

For negative indices, we are still in a position to define corresponding Sobolev classes. The object collected in $\mathcal{S}(a, M)$ *for* $a < 0$ *need no longer be square integrable functions, but are those tempered distributions for which*

$$\int (1+|u|^2)^a |\mathcal{F}f(-u)|^2\, du \leq M \tag{2.3.4}$$

holds.

Next, let us recall the following definition:

Definition 2.3.2. *For* $k \in \mathbb{R}$*, a kernel* K *is called a* k-th order kernel*, if for all integers* $1 \leq m < k$*,*

$$\int x^m K(x)\, dx = 0 \tag{2.3.5}$$

and moreover,

$$\int |x|^k |\,\mathrm{K}(x)|\, dx < \infty. \tag{2.3.6}$$

Formula (2.3.5) is equivalent to stating that the derivatives $(\mathcal{F}\,\mathrm{K})^{(m)}(0)$ vanish for $1 \leq m < k$.

Rate results under global regularity assumptions

We start by providing rate results under global regularity assumptions on the test function f and on μ, measured in a Sobolev sense.

Let us first have a look at the approximation error which results from smoothing with a kernel function K:

Lemma 2.3.3. *Assume that for some real valued* s *and some positive constant* M_f*, we have* $f \in \mathcal{S}(s, M_f)$*. Assume, moreover, that for some* $a > -s$*, and some positive constant* M_μ*,* $\mu \in \mathcal{S}(a, M_\mu)$*. Let* K *be chosen such that either* K *is the sinc kernel or* K *has order* $a+s$ *and* $\mathcal{F}\,\mathrm{K}$ *is Hölder regular with index* $a+s$ *and constant* L_{K}*. Then we can estimate*

$$\left| \int f(x)\mu(\,dx) - \int f(x)(\mathrm{K}_h * \mu)(x)\, dx \right|^2 \leq \frac{C_B}{(2\pi)^2} h^{2a+2s} =: b_h \tag{2.3.7}$$

with constant

$$C_B = \left(2\pi^{-s-a} + \left(\frac{L_K}{\langle a+s \rangle !} \right) \right)^2 M_f M_\mu. \tag{2.3.8}$$

Next, we have the following bound on the error in the model:

Lemma 2.3.4. *Assume that $\mathcal{F}\mathrm{K}$ is supported on $[-\pi, \pi]$. Assume, moreover, that $f \in \mathcal{S}(s, M_f)$ and that for positive constants C_φ and c_φ,*

$$\forall u \in \mathbb{R} : |\varphi_\Delta(u)| \geq C_\varphi (1 + |u|^2)^{-\frac{\Delta\beta}{2}} \exp\left(-\Delta c_\varphi |u|^\rho\right). \tag{2.3.9}$$

Let

$$\sigma_h^2 := \frac{C_1}{2\pi^2} \int |\mathcal{F}\mathrm{K}(hu)|^2 \left| \frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)} \right|^2 du \wedge \frac{C_2}{2\pi^2} \left(\int |\mathcal{F}\mathrm{K}(hu)| \left| \frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)} \right| du \right)^2 \Bigg\}. \tag{2.3.10}$$

Then we have

$$\sigma_h^2 \leq \frac{C_\varphi M_f}{2\pi^2} \Bigg\{ C_1 \sup_{|u| \leq \frac{\pi}{h}} (1 + |u|)^{2\Delta\beta - 2s} \exp\left(2\Delta c_\varphi |u|^\rho\right) \wedge C_2 \int_{\{|u| \leq \frac{\pi}{h}\}} (1 + |u|)^{2\Delta\beta - 2s} \exp\left(2\Delta c_\varphi |u|^\rho\right) du \Bigg\} =: \upsilon_h. \tag{2.3.11}$$

Now, let us introduce the following abstract nonparametric classes of signed measures:

Definition 2.3.5. *Let $\mathcal{M} := \mathcal{M}(\bar{C}_1, \bar{C}_2, C_\varphi, c_\varphi, \beta, \rho, a, M_\mu)$ be the collection of finite signed measures μ, such that the following holds true:*

(i) There is a Lévy process X such that the assumptions (A1)-(A4) are satisfied and for the corresponding Lévy measure ν, $\mu(\,dx) = x\nu(\,dx)$.

(ii) For

$$\varphi(u) := \exp\left(\int \frac{e^{iux} - 1}{x} \mu(\,dx) \right), \tag{2.3.12}$$

we have that

$$\forall u \in \mathbb{R} : \ |\varphi(u)| \geq C_\varphi (1 + |u|^2)^{-\frac{\beta}{2}} e^{-c_\varphi |u|^\rho}. \tag{2.3.13}$$

(iii) For C_1 and C_2 defined as in (2.2.21) and (2.2.22), we have $C_1 \leq \bar{C}_1$ and $C_2 \leq \bar{C}_2$.

(iv) μ is contained in the Sobolev class $\mathcal{S}(a, M_\mu)$.

Let $\mathbb{P}_\mu$ be the infinitely divisible law with characteristic function φ defined by (2.3.12) and $\mathbb{E}_\mu$ the expectation with respect to $\mathbb{P}_\mu$.

We can now provide rates of convergence, uniformly over those nonparametric classes:

Theorem 2.3.6. *Assume that $f \in \mathcal{S}(s, M_f)$. Consider the nonparametric class $\mathcal{M} := \mathcal{M}(\bar{C}_1, \bar{C}_2, C_\varphi, c_\varphi, \beta, \rho, a, M_\mu)$ with $a > -s$. For $h > 0$, let $\widehat{\theta}_{\Delta,h,n}$ be the kernel estimator defined according to Definition 2.2.7. Assume that the conditions on the kernel function which are summarised in Lemma 2.3.3 and Lemma*

2.3.4 are met. Let b_h and v_h be defined as in Lemma 2.3.3 and Lemma 2.3.4. Then, selecting $h^* = h^*_{\Delta,n}$ as the minimiser of $b_h + T^{-1}v_h$, we derive that

$$\sup_{\mu\in\mathcal{M}} \mathbb{E}_\mu \left[\left| \theta - \widehat{\theta}_{h^*,\Delta,n} \right|^2 \right] = O\left(r_{\Delta,n}\right) \tag{2.3.14}$$

with $(r_{\Delta,n})$ denoting the sequences which are summarised in the following table:

	$\overline{C}_1 < \infty$		$\overline{C}_1 = \infty$	
$\rho = 0$	$s \geq \Delta\beta$	T^{-1}	$s \geq \Delta\beta + \frac{1}{2}$	T^{-1}
	$s = \Delta\beta$	T^{-1}	$s = \Delta\beta + \frac{1}{2}$	$(\log T)T^{-1}$
	$s < \Delta\beta$	$T^{-\frac{2a+2s}{2a+2\Delta\beta}}$	$s < \Delta\beta + \frac{1}{2}$	$n^{-\frac{2a+2s}{2a+2\Delta\beta+1}}$
$\rho > 0$		$(\frac{\log T}{\Delta})^{-\frac{2a+2s}{\rho}}$		$(\frac{\log T}{\Delta})^{-\frac{2a+2s}{\rho}}$

Let us compare the rates of convergence derived above to the rates of convergence which can be obtained for the estimator $\widehat{\theta}_{\Delta,n}$, which is defined without an additional smoothing procedure.

Theorem 2.3.7. *Let $f \in \mathcal{S}(s, M_f)$ for some $s > \frac{1}{2}$. Consider the nonparametric class $\mathcal{M} := \mathcal{M}(\bar{C}_1, \bar{C}_2, C_\varphi, c_\varphi, \beta, \rho, a, M_\mu)$. Let $\widehat{\theta}_{\Delta,n}$ be defined by (2.2.19). Then we find that*

$$\sup_{\mu\in\mathcal{M}} \mathbb{E}_\mu \left[\left| \theta - \widehat{\theta}_{\Delta,n} \right|^2 \right] = O\left(r_{\Delta,n}\right), \tag{2.3.15}$$

with $(r_{\Delta,n})$ collected in the following table:

	$\bar{C}_1 < \infty$		$\bar{C}_1 = \infty$	
$\rho = 0$	$s > \Delta\beta$	T^{-1}	$s > \Delta\beta + \frac{1}{2}$	T^{-1}
	$s = \Delta\beta$	T^{-1}	$s = \Delta\beta + \frac{1}{2}$	$(\log T)T^{-1}$
	$s < \Delta\beta$	$T^{-\frac{2s}{2\Delta\beta}}$	$s < \Delta\beta + \frac{1}{2}$	$T^{-\frac{(2s-1)}{2\Delta\beta}}$
$\rho > 0$		$\left(\frac{\log T}{\Delta}\right)^{-\frac{2s}{\rho}}$		$\left(\frac{\log T}{\Delta}\right)^{-\frac{2a+2s}{\rho}}$

It is interesting to note that we basically recover, in Theorem 2.3.7 for $\bar{C}_1 = \infty$ the rates of convergence which have been derived by Neumann and Reiß in [56] for $\Delta = 1$ and the dependence on Δ which was found by Kappus and Reiß in [38]. This confirms the analogy between our constructive estimator and the minimum distance estimator.

However, in [56], a logarithmic loss was found in the rate for processes with polynomially decaying Fourier transform, which does not occur in the present setting. We can thus not confirm the authors' suspicion that the logarithmic gap between upper and lower bound was due to a suboptimal lower bound.

Moreover, the parametric rate results should be compared to the results recently found by Nickl and Reiß [58].

Let us briefly comment on the meaning of the constants C_1 and C_2. The definition of C_1 involves

$$\int |\Psi'(u)|^2 \,\mathrm{d}u = \int |\mathcal{F}(x\nu(\,\mathrm{d}x))(u)|^2 \,\mathrm{d}u = \int |\mathcal{F}\mu(u)|^2 \,\mathrm{d}u \tag{2.3.16}$$

as well as

$$\int |\Psi''(u)| \,\mathrm{d}u = \int |\mathcal{F}(x^2\nu(\,\mathrm{d}x))(u)| \,\mathrm{d}u = \int |(\mathcal{F}(x\mu(\,\mathrm{d}x))(u)| \,\mathrm{d}u. \tag{2.3.17}$$

These quantities are finite if $\mu(\,\mathrm{d}x)$ has a square integrable Lebegue density and $x\mu(\,\mathrm{d}x)$ is integrable, with integrable Fourier transform, which basically tells us that ν has a Lebesgue density which is smooth away from the origin and has moderate growth near the origin.

Examples 2.3.8.

(i) *For Gamma processes, a direct calculation of the Fourier transforms gives* $|\mathcal{F}\mu(u)| \sim |u|^{-1},\ |u| \to \infty$ *and* $|\mathcal{F}(x\mu(\,dx))(u)| \sim |u|^{-2},\ |u| \to \infty$*, which readily implies finiteness of* C_1.

(ii) *If* $\nu(\{y\}) \neq 0$ *for some* $y \in \mathbb{R}\backslash\{0\}$*, the Fourier transforms of* μ *and* $x\mu(\,dx)$ *do not tend to zero as* $|u|$ *tends to infinity, so* C_1 *cannot be finite.*

(iii) *For tempered stable distributions, we have* $\mathcal{F}\mu(u) \sim |u|^{-1+\alpha}$ *and* $\mathcal{F}(x\mu(\,dx)) \sim |u|^{-2+\alpha}$*. Consequently,* C_1 *is finite for* $\alpha < 1/2$ *and infinite else.*

Notice that the constant C_2 is finite, in any case, since we have $\|\mathcal{F}\mu\|_\infty \leq |\mu|(\mathbb{R})$ and $\|\mathcal{F}(x\mu(\,\mathrm{d}x))\|_\infty < (x\mu(\,\mathrm{d}x))(\mathbb{R})$, and these quantities are finite thanks to the assumptions *(A2)* and *(A4)*.

Let us compare the results which are found in Theorem 2.3.6 to the rates which are derived in Theorem 2.3.7.

As usual in deconvolution problems, the faster the absolute value of φ decays, the worse are the rates of convergence to be obtained.

Still, when considering density deconvolution problems, the rates get better if the underlying probability density is sufficiently regular. The same is true, in a sense, in the present setting, since the rates improve as a increases.

However, very much unlike in classical density deconvolution, when turning to the Lévy model, we have to beware of the fact that a, β and ρ are by no means independent of each other.

If f has integrable Fourier transform, the results of Theorem 2.3.7 tell us that the rates of convergence which are obtained without any smoothing procedure are parametric if $s \geq \Delta\beta$. For example, we will automatically attain the parametric rate when we are in the compound Poisson setting, that is, when $\beta = 0$ or when we sample at a high enough frequency.

On the other hand, it was seen in Section 1.2 that for $\beta > 0$, the Lévy density will have a Gamma-like behaviour around the origin, so the best we can hope for is that μ has a density g with Fourier transform decaying as $|\mathcal{F}g(u)| \sim |u|^{-1}$, which gives $a < \frac{1}{2}$. Moreover, the larger β becomes, the larger will M_μ be.

In case that $|\varphi|$ has exponential decay, $\mathcal{F}\mu$ will behave as $|u|^{-\rho}$, making $a < \frac{1}{2} - \rho$. Consequently, the possible gain in the rates which results from applying some smoothing procedure rather than simply considering $\widehat{\theta}_{\Delta,n}$ is small, in any case.

One could describe this phenomenon as "ill-posedness coming from two sides". A fast decay of φ does not only, in it self, lead to slow rates of convergence, but also forbids the existence of a relatively smooth underlying measure μ.

Of course, this problem is not particular to the estimation of linear functionals, but is also found when estimating g in an L^2-sense.

We conclude that, for smooth test functions f which do not vanish at zero, it will often make more sense in applications to consider $\widehat{\theta}_{\Delta,n}$ instead of working with some kernel smoothing.

The situation is different, however, when we consider test functions or general distributions f which are bounded away from the origin.

Rate results under local regularity assumptions

Let us now consider the rate results which can be obtained under local regularity assumptions on μ, typically measured in a Hölder sense, when applying a kernel function which decays fast enough.

We can give the following bound on the approximation error:

Lemma 2.3.9. *Let f be compactly supported with $\operatorname{supp}(f) := [a,b] \subseteq \mathbb{R}\setminus\{0\}$ and assume that for some $s \in \mathbb{Z}$ and some positive integer C_f,*

$$\forall u \in \mathbb{R} : \ |\mathcal{F}f(u)| \le C_f(1+|u|)^{-s} \tag{2.3.18}$$

holds true.

Assume, moreover, that for some bounded open set $D := (d_1, d_2) \supseteq [a,b]$, μ possesses locally a Lebesgue density $g_D \in \mathcal{H}_D(a,R,L)$ with $a > -s$.

Let the order of K be $a+s$. Assume that K is $-s \vee 0$-times continuously differentiable. Assume, moreover, that there is a positive constant C_K such that for any nonnegative integer $m \le (0 \vee -s)$,

$$\forall z \in \mathbb{R} : |\,\mathrm{K}^{(m)}(z)| \le C_K(1+|z|)^{-(a+s)-m-1} \tag{2.3.19}$$

holds true.

Then we can give the following bound on the approximation error:

$$\left| \int f(x)\mu(\,dx) - \int f(x)(\mathrm{K}_h * \mu)(x)\,dx \right|^2 \le C_B h^{2a+2s} \tag{2.3.20}$$

with some positive constant C_B depending on $C_K, a - d_1 \vee d_2 - b$, L and R.

The following result is in analogy with Lemma 2.3.4. However, we need to pay attention on the fact, that the definition of the smoothness parameter s is now slightly different.

Lemma 2.3.10. *In the situation of the preceding lemma, assume that for positive constants C_φ and c_φ, we have*

$$\forall u \in \mathbb{R}: \; C_\varphi (1+|u|^2)^{-\frac{\Delta\beta}{2}} e^{-\Delta c_\varphi |u|^\rho} \leq |\varphi_\Delta(u)|. \tag{2.3.21}$$

Assume, moreover, that $\mathcal{F}\mathrm{K}$ is supported on $[-\pi,\pi]$. Then, with σ_h^2 defined as in Lemma 2.3.4, we have

$$\begin{aligned}\sigma_h^2 \leq \frac{C_f^2 C_\varphi^2}{2\pi^2} \Bigg\{ & C_1 \int\limits_{\{|u|\leq \frac{\pi}{h}\}} (1+|u|)^{2\Delta\beta - 2s} \exp\left(2\Delta c_\varphi |u|^\rho\right)\, du \\ & \wedge C_2 \left(\int\limits_{\{|u|\leq \frac{\pi}{h}\}} (1+|u|)^{\Delta\beta - s} \exp\left(\Delta c_\varphi |u|^\rho\right)\, du \right)^2 \Bigg\} =: v_h. \end{aligned} \tag{2.3.22}$$

We consider the following classes of locally Hölder regular measures:

Definition 2.3.11. *Let $\mathcal{M} := \mathcal{M}(\bar{C}_1, \bar{C}_2, C_\varphi, c_\varphi, \beta, \rho, a, D, L, R)$ be the collection of finite signed measures μ, such that the following holds: The items (i)-(iii) from Definition 2.3.5 are true and*

(iv) $D = (d_1, d_2)$ is a bounded open interval, $\mu|_D$ possesses a Lebesgue density g_D which is Hölder-regular in the sense that $g_D \in \mathcal{H}_D(a, L, R)$.

The rate results which can be derived from Lemma 2.3.9 and Lemma 2.3.10 are summarised in the following theorem:

Theorem 2.3.12. *Let the assumptions of Lemma 2.3.9 and Lemma 2.3.10 be satisfied. Consider the nonparametric class $\mathcal{M} = \mathcal{M}(\bar{C}_1, \bar{C}_2, C_\varphi, c_\varphi, \beta, \rho, a, D, L, R)$ defined in Definition 2.3.11. Let h^* be selected as the minimiser of $b_h + T^{-1} v_h$. Then we find that*

$$\sup_{\mu \in \mathcal{M}} \mathbb{E}_\mu \left[\left| \theta - \widehat{\theta}_{\Delta, h^*, n} \right|^2 \right] = O\left(r_{\Delta,n}\right) \tag{2.3.23}$$

with the rates $r_{\Delta,n}$ collected in the following table:

	$\overline{C}_1 < \infty$		$\overline{C}_1 = \infty$	
$\rho = 0$	$s > \Delta\beta + \frac{1}{2}$	T^{-1}	$s > \Delta\beta + 1$	T^{-1}
	$s = \Delta\beta + \frac{1}{2}$	$(\log T)\, T^{-1}$	$s = \Delta\beta + 1$	$(\log T) T^{-1}$
	$s < \Delta\beta + \frac{1}{2}$	$T^{-\frac{2s+2a}{2\Delta\beta + 2a + 1}}$	$s < \Delta\beta + 1$	$T^{-\frac{2a+2s}{2a+2\Delta\beta+2}}$
$\rho > 0$		$\left(\frac{\log T}{\Delta}\right)^{-\frac{2a+2s}{\rho}}$		$\left(\frac{\log T}{\Delta}\right)^{-\frac{2a+2s}{\rho}}$

Examples 2.3.13.

(i) *Consider point estimation. If $\rho = 0$, the measures collected in the abstract smoothness class $\mathcal{M}$ correspond to Lévy processes having finite or moderate activity of small jumps and for which the Lévy measure is locally Hölder regular. Typical examples are Gamma-processes or Compound Poisson processes with smooth jump density. In this cases, we have $s = 0$ and we recover the rate of convergence $O\left(T^{-\frac{2a}{2a+2\Delta\beta+1}}\right)$ which is classical for point estimation in density deconvolution.*

It should not come as a surprise that we recover in the continuous limit, that is, for Δ close to zero, the rate $O\left(T^{-\frac{2a}{2a+1}}\right)$, which is classical for density estimation with pointwise loss.

(ii) *When estimating the k-th derivative at some point $y \in \mathbb{R}\setminus\{0\}$, we have $k = -s$. The rates of convergence $O\left(T^{-\frac{2a-2k}{2a+2\Delta\beta+1}}\right)$ match, again, the rates which are known from density deconvolution (see [26]) and one recovers, in the continuous limit, the rates which are standard in density estimation.*

(iii) *Next, consider estimation of integrals of the form $\int 1(A)\nu(\,dx)$ for compact sets A bounded away from the origin. Clearly, we have in this case, $s = 1$. For $\rho = 0$ and small values of β, namely $\beta < \frac{1}{2}$, we end up with the parametric rate $O\left(T^{-1}\right)$. For values $\beta > \frac{1}{2}$, we have the order $O\left(T^{-\frac{2a+2}{2a+2\beta+1}}\right)$*

2.4 Lower bound results

In the last section, we have derived rates of convergence for estimating linear functionals over nonparametric classes of Lévy processes.

It is the aim of the present section to show that the convergence rates thus obtained cannot be improved, in other words, to provide minimax lower bounds.

However we do not give a fully rigorous discussion for arbitrary choices of f, but content ourselves with considering the particularly interesting cases of point estimation and estimating integrals.

Recall the following definition:

Definition 2.4.1. *Let a statistical model $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ be given. Let d be some semi-distance on the parameter space Θ and let $\mathcal{M} \subseteq \Theta$ be some subset of the parameter space. The* minimax risk *r^* over $\mathcal{M}$ is defined to be*

$$r^* := \inf_{\bar\theta} \sup_{\theta\in\mathcal{M}} \mathbb{E}_\theta\left[\, d^2(\bar\theta, \theta)\right], \tag{2.4.1}$$

where the infimum is taken over all estimators $\bar\theta : \mathcal{X} \to \Theta$.

From a minimax point of view, it is desirable that for some prescribed subset $\mathcal{M}$ of the parameter space, which is assumed to contain the true parameter θ, the maximal risk of an estimator over the collection $\mathcal{M}$ is close to the minimax risk.

The performance of an estimator $\widehat{\theta}$ is thus measured by the ratio

$$\frac{\sup_{\theta\in\mathcal{M}} \mathbb{E}_\theta \left[d^2(\widehat{\theta},\theta)\right]}{r^*}. \tag{2.4.2}$$

From an asymptotic point of view, one should wish that for a sequence of statistical experiments, the ratio is uniformly bounded. This leads to the following definition:

Definition 2.4.2. *Let $(\mathcal{X}^n, \mathcal{A}^n, (\mathbb{P}^n_\theta)_{\theta\in\Theta})$ be a sequence of statistical models. Let $\mathcal{M} \subseteq \Theta$ and let r^*_n denote the minimax risk over $\mathcal{M}$ in the n-th model. A sequence $(r_n)_{n\in\mathbb{N}}$ is called an* optimal rate of convergence over $\mathcal{M}$ *if the following holds true:*

For some constant $C < \infty$,

$$\limsup_{n\to\infty} \frac{r^*_n}{r^2_n} \leq C \tag{2.4.3}$$

and moreover, for some constant $c > 0$,

$$\liminf_{n\to\infty} \frac{r^*_n}{r^2_n} \geq c. \tag{2.4.4}$$

A sequence of estimators $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ is called asymptotically minimax optimal, *if for the optimal rate of convergence $(r_n)_{n\in\mathbb{N}}$ and some positive constant C',*

$$\forall n \in \mathbb{N}: \sup_{\theta\in\mathcal{M}} \mathbb{E}_\theta \left[d^2(\widehat{\theta}_n,\theta)\right] \leq C' r^2_n \tag{2.4.5}$$

holds.

In the preceding section, we have shown that for some positive constant C,

$$\forall n \in \mathbb{N}: \sup_{\mu\in\mathcal{M}} \mathbb{E}_\mu \left[|\theta - \widehat{\theta}_{\Delta,h^*,n}|^2\right] \leq C r_{\Delta,n} \tag{2.4.6}$$

holds true, for classes $\mathcal{M}$ of measures which are locally Hölder smooth.

To see that the rates of convergence thus obtained cannot be improved, we have to check that

$$\liminf_{n\to\infty} \inf_{\bar{\theta}_n} \sup_{\mu\in\mathcal{M}} \mathbb{E}_\mu \left[|\theta - \widehat{\theta}_{\Delta,h^*,n}|^2\right] r^{-1}_{\Delta,n} > 0. \tag{2.4.7}$$

This is done by using the fact that *estimating* a parameter within some prescribed collection $\mathcal{M}$ is always more complicated then *testing* between two or finitely many alternatives in $\mathcal{M}$. A detailed discussion and the proof of the following lemma can be found in Chapter 2 in [70].

Lemma 2.4.3. *Let a sequence of statistical models be given. Let d be some semi-distance defined on Θ. Let $\mathcal{M} \subseteq \Theta$ be some subset of the parameter space.*

Let $(\theta_{0,n})_{n\in\mathbb{N}}$ and $(\theta_{1,n})_{n\in\mathbb{N}}$ be sequences in $\mathcal{M}$ such that for some positive constant c,

$$\forall n \in \mathbb{N}: \quad d^2(\theta_{0,n}, \theta_{1,n}) \geq c r_n^2. \tag{2.4.8}$$

Then the following estimate holds for the minimax risk:

$$\liminf_{n\to\infty} \inf_{\bar{\theta}_n} \sup_{\theta\in\mathcal{M}} \mathbb{E}_\theta \left[d^2(\theta, \bar{\theta}_n) \right] r_n^{-2} \geq \liminf_{n\to\infty} \inf_{\psi_n} \sup_{j=0,1} \mathbb{P}_{\theta_{j,n}} \left(\{\psi_n \neq j\}\right), \tag{2.4.9}$$

where the infimum is taken over all tests $\psi_n : \mathcal{X}^n \to \{0,1\}$.

This tells us that what has to be done to derive lower bounds of the form (2.4.7) is to construct a sequence of alternatives within the prescribed smoothness class such that the distances decay with rate r_n and that the corresponding distributions do not separate in the sense that

$$\liminf_{n\to\infty} \inf_{\psi_n} \sup_{j=0,1} \mathbb{P}_{\theta_{j,n}} \left(\{\psi_n \neq j\}\right) > 0 \tag{2.4.10}$$

holds true.

To do this, we need the following definition and Theorem (again, we refer to [70] for a detailed discussion and for the proof):

Definition 2.4.4. *Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures defined on a common measurable space. Let λ be a dominating measure. The χ^2-*distance *is*

$$\chi^2(\mathbb{P},\mathbb{Q}) := \begin{cases} \int \left(\frac{d\mathbb{P}}{d\mathbb{Q}} - 1\right)^2 \, d\lambda, & \text{if } \mathbb{P} \ll \mathbb{Q} \\ \infty, & \text{else.} \end{cases} \tag{2.4.11}$$

Theorem 2.4.5. *In the situation of Lemma 2.4.3, the inequality (2.4.10) holds true if the following holds for the χ^2-distances of the distributions corresponding to $\theta_{0,n}$ and $\theta_{1,n}$:*

$$\limsup_{n\to\infty} \chi^2(\mathbb{P}_{0,n}, \mathbb{P}_{1,n}) = O\left(n^{-1}\right). \tag{2.4.12}$$

We are now ready to formulate the lower bound results for estimating integrals and for point estimation in the Lévy model. For sake of simplicity, we content ourselves with the case of polynomially decaying Fourier transforms.

Theorem 2.4.6. *Let $\mathcal{M} = \mathcal{M}(\bar{C}_1, \bar{C}_2, C_\varphi, c_\varphi, \beta, \rho, a, D, L, R)$ be the collection of signed measures defined as in Definition 2.3.11. Let $\bar{C}_1 < \infty$ and $\rho = 0$. Let $[a,b] \subseteq D$ and $f(x) = x^{-1} 1_{[a,b]}(x)$. Then we have for $\theta = \int f(x)\mu(\,dx) = \nu([a,b])$:*

$$\liminf_{n\to\infty} \inf_{\bar{\theta}_n} \sup_{\mu\in\mathcal{M}} \mathbb{E}_\mu \left[\left|\bar{\theta}_{\Delta,n} - \theta\right|^2 \right] T^{-\frac{2a+2}{2a+2\Delta\beta+1}} > 0, \tag{2.4.13}$$

where the infimum is taken over all estimators based on n observations of the process at time points $\Delta, \cdots, 2\Delta n$.

Proof. For notational convenience, we may assume, without loss of generality, that $D = (\pi, 4\pi)$ and $[a,b] = [2\pi, 3\pi]$. Moreover, since Δ is assumed to be fixed, we can drop the dependence on Δ and write, henceforth, β instead of $\Delta\beta$ and

n instead of T. Finally, we can assume, without loss of generality, that $\beta \geq \frac{1}{2}$ since a simple parametric argument shows that the rate cannot be better then the parametric rate.

We start by considering some symmetric bilateral Gamma distribution $\mathbb{P}_0$. A bilateral Gamma distribution is the convolution of a $\Gamma(\beta^-, \lambda^-)$-distribution supported on the negative half axes and a $\Gamma(\beta^+, \lambda^+)$-distribution supported on the positive half axes, see Küchler and Tappe [43].

We let $\beta^+ = \beta^- = \frac{\beta}{2}$, where β is the parameter appearing in the definition of $\mathcal{M}$. Moreover, we let $\lambda^+ = \lambda^- = \lambda$ for some $\lambda > 0$ to be chosen.

The Lévy measure ν_0 corresponding to $\mathbb{P}_0$ has Lebesgue density

$$\eta_0 = \frac{\beta}{2}|x|^{-1}e^{-\lambda|x|}. \tag{2.4.14}$$

The characteristic function φ_0 of $\mathbb{P}_0$ is given by

$$\varphi_0(u) = \left(1 + \left(\frac{u}{\lambda}\right)^2\right)^{-\frac{\beta}{2}}. \tag{2.4.15}$$

Let $\mu_0 := x\nu_0(\,\mathrm{d}x)$ and let g_0 be the Lebesgue density of μ_0. We have to convince ourselves that μ_0 belongs to the nonparametric class $\mathcal{M}$.

It follows readily from (2.4.15) that

$$|\varphi_0(u)| \geq C_\varphi(1 + |u|^2)^{-\beta/2} \tag{2.4.16}$$

holds true, provided that λ is chosen large enough.

Next, we notice that by formula (2.4.14), we have

$$\mathcal{F}\left(x\mu_0(\,\mathrm{d}x)\right)(u) = \beta \int_0^\infty e^{iux} x e^{-\lambda x}\,\mathrm{d}x = \frac{\beta}{(\lambda - iu)^2}. \tag{2.4.17}$$

and

$$\mathcal{F}\mu_0(u) = \frac{\beta}{2}\int_0^\infty e^{iux}e^{-\lambda x}\,\mathrm{d}x - \frac{\beta}{2}\int_{-\infty}^0 e^{iux}e^{\lambda x}\,\mathrm{d}x = \frac{\beta i u}{\lambda^2 + u^2}. \tag{2.4.18}$$

From this we conclude that we have

$$C_1 = C\left(\int |\mathcal{F}(x\mu_0(x))(u)|\,\mathrm{d}x + 2\int |(\mathcal{F}\mu_0)(u)|^2\,\mathrm{d}u\right) < C_1' \tag{2.4.19}$$

for some $C_1' < \bar{C}_1$, provided that λ is chosen large enough.

Finally, it follows immediately from formula (2.4.14) that for large enough values of λ, we have $g_0\big|_D \in \mathcal{H}_D(a, L', R')$ for some $R' < R$ and $L' < L$. From this we conclude that $\mathbb{P}_0$ belongs to $\mathcal{M}$, for λ chosen large enough.

To construct an alternative within the prescribed smoothness class, we add to η_0 a small perturbation at the boundary of $\mathrm{supp}(f)$. Let h_0 be some nontrivial function which is supported on $[-\frac{\pi}{2}, \frac{\pi}{2}]$, nonnegative on $[0, \frac{\pi}{2}]$ and for which $h_0(-x) = -h_0(x)$ holds. Moreover, let h_0 be chosen such that $x \mapsto (x+2\pi)h_0(x)$ belongs to $\mathcal{H}_{[-\frac{\pi}{2},\frac{\pi}{2}]}(a, L-L', R-R')$.

For $K \geq 1$, let $h_K := K^{-a}h_0(Kx)$ and let

$$\eta_K(x) := \eta_0(x) + \gamma h_K(x-2\pi), \tag{2.4.20}$$

for some positive constant γ to be chosen.

We have to check that the signed measure μ_K which has Lebesgue density $g_K(x) = x\eta_K(x)$ belongs to $\mathcal{M}$:

First of all, η_K is clearly the Lebesgue density of a Lévy measure, provided that γ is chosen small enough to ensure that η_K is nonnegative.

Local Hölder regularity of g_K follows readily from Hölder regularity of the original density g_0 and from the fact that the perturbation is Hölder-regular by construction.

Next, since we have $C_1 \leq C_1' < \bar{C}_1$, we can ensure that

$$C\left(2\|g_K\|_{\mathrm{L}^2}^2 + \|xg_K(x)\|_{\mathrm{L}^1}\right) \leq \bar{C}_1$$

holds true by choosing γ small enough.

Finally, we have to check that the Fourier transform φ_K of the infinitely divisible distribution corresponding to g_K has the same decay behaviour as φ_0. The definition of g_K yields

$$\begin{aligned}\varphi_K(u) &= \exp\left(\int \frac{(e^{iux}-1)}{x}(g_0(x) + \gamma x h_K(x-2\pi))\,\mathrm{d}x\right)\\ &= \varphi_0(u)\exp\left(\gamma\int (e^{iux}-1)h_K(x-2\pi)\,\mathrm{d}x\right).\end{aligned} \tag{2.4.21}$$

Since h_K is antisymmetric by assumption, we find that

$$\int 1 h_K(x-2\pi)\,\mathrm{d}x = 0 \tag{2.4.22}$$

and hence

$$\begin{aligned}&\varphi_0(u)\exp\left(\gamma\int (e^{iux}-1)h_K(x-2\pi)\,\mathrm{d}x\right)\\ &= \varphi_0(u)\exp\left(\gamma\int e^{iux}h_K(x-2\pi)\,\mathrm{d}x\right) = \varphi_0(u)\exp\left(\gamma e^{iu2\pi}\mathcal{F}h_K(u)\right).\end{aligned} \tag{2.4.23}$$

Since $|\mathcal{F}h_K(u)|$ is uniformly bounded, we conclude that φ_K has the same decay behaviour as φ_0.

We have thus shown that $\mu_K \in \mathcal{M}$.

The loss between the original density g_0 and the perturbed version g_K can be calculated as follows:

$$\left|\int f(x)(g_K(x) - g_0(x))\,\mathrm{d}x\right| = \left|\int\limits_{[2\pi,3\pi]} x^{-1}(g_K(x) - g_0(x))\,\mathrm{d}x\right|$$

$$
\begin{aligned}
&= \left| \int\limits_{[2\pi,3\pi]} (\eta_K(x) - \eta_0(x))\,\mathrm{d}x \right| = \gamma \left| \int\limits_{[2\pi,3\pi]} h_K(x-2\pi)\,\mathrm{d}x \right| \\
&= \gamma \left| \int\limits_{[0,\frac{\pi}{2}]} K^{-a} h_0(Kx)\,\mathrm{d}x \right| = \gamma \left| \int\limits_{[0,\frac{\pi}{2}]} h_0(x)\,\mathrm{d}x \right| K^{-a-1}. \qquad (2.4.24)
\end{aligned}
$$

Next, we have to calculate the χ^2-distance between $\mathbb{P}_0$ and $\mathbb{P}_K$. Let f_0 and f_K denote the probability densities corresponding to $\mathbb{P}_0$ and $\mathbb{P}_K$, respectively. We apply here the strategy which has been proposed in [56] and use the fact that for some positive constant c, we have $f_0(x) \geq c^{-1} e^{-\lambda|x|}$. From this we derive that

$$
\begin{aligned}
&\chi^2(\mathbb{P}_0, \mathbb{P}_K) = \int \frac{(f_0(x) - f_K(x))^2}{f_0(x)}\,\mathrm{d}x \\
&\leq c \int \left(e^{\frac{\lambda}{2}|x|} (f_0(x) - f_K(x)) \right)^2 \mathrm{d}x \\
&\leq c \int \left(e^{\frac{\lambda}{2}x} (f_0(x) - f_K(x)) \right)^2 \mathrm{d}x + c \int \left(e^{-\frac{\lambda}{2}x} (f_0(x) - f_K) \right)^2 \mathrm{d}x. \qquad (2.4.25)
\end{aligned}
$$

Passing to the Fourier domain, we find that

$$
\begin{aligned}
& c \int \left(e^{\frac{\lambda}{2}x} (f_0(x) - f_K(x)) \right)^2 \mathrm{d}x + c \int \left(e^{-\frac{\lambda}{2}x} (f_0(x) - f_K) \right)^2 \mathrm{d}x \\
= \;& \frac{c}{2\pi} \int \left(\mathcal{F}(e^{\frac{\lambda}{2}x} f_0(x))(u) - \mathcal{F}(e^{\frac{\lambda}{2}x} f_K(x))(u) \right)^2 \mathrm{d}u \\
+ \;& \frac{c}{2\pi} \int \left(\mathcal{F}(e^{-\frac{\lambda}{2}x} f_0(x))(u) - \mathcal{F}(e^{-\frac{\lambda}{2}x} f_K(x))(u) \right)^2 \mathrm{d}u. \qquad (2.4.26)
\end{aligned}
$$

Now, Theorem 25.17 in Sato [63] allows to calculate

$$
\begin{aligned}
& \mathcal{F}\left(e^{\frac{\lambda}{2}x} f_0(x) \right)(u) = \int e^{\frac{\lambda}{2}x} f_0(x) e^{iux}\,\mathrm{d}x = \int f_0(x) e^{i\left(u - \frac{i\lambda}{2}\right)x}\,\mathrm{d}x \\
= \;& \mathcal{F} f_0\,(u - i\lambda/2) = \left(1 + (u/\lambda - i/2)^2 \right)^{-\frac{\beta}{2}} \qquad (2.4.27)
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathcal{F}\left(e^{\frac{\lambda}{2}x} f_K(x) \right)(u) = \mathcal{F} f_K\,(u - i\lambda/2) \\
= \;& \exp\left(\int \frac{e^{i(u - \frac{\lambda}{2}i)x} - 1}{x} (g_0(x) + \gamma x h_K(x - 2\pi))\,\mathrm{d}x \right) \\
= \;& \mathcal{F} f_0\,(u - i\lambda/2) \exp\left(\gamma \int e^{i\left(u - \frac{\lambda}{2}i\right)x} h_K(x - 2\pi)\,\mathrm{d}x \right) \\
= \;& \left(1 + (u/\lambda - i/2)^2 \right)^{-\frac{\beta}{2}} \exp\left(\gamma \int e^{i\left(u - \frac{\lambda}{2}i\right)x} h_K(x - 2\pi)\,\mathrm{d}x \right). \qquad (2.4.28)
\end{aligned}
$$

The same reasoning gives

$$\mathcal{F}(e^{-\frac{\lambda}{2}x}f_0(x))(u) = \left(1 + (u/\lambda + i/2)^2\right)^{-\frac{\beta}{2}} \tag{2.4.29}$$

and

$$\begin{aligned} &\mathcal{F}\left(e^{-\frac{\lambda}{2}x}f_K(x)\right)(u) \\ = \quad & \left(1 + (u/\lambda + i/2)^2\right)^{-\frac{\beta}{2}} \exp\left(\gamma \int e^{i\left(u+\frac{\lambda}{2}i\right)x} h_K(x-2\pi)\,\mathrm{d}x\right). \end{aligned} \tag{2.4.30}$$

We have thus shown that

$$\begin{aligned} &\chi^2\left(\mathbb{P}_0, \mathbb{P}_K\right) \\ \leq \quad & c\int \left|\frac{3}{4} + \frac{u^2}{\lambda^2}\right|^{-\beta} \left|1 - \exp\left(\gamma \int e^{i\left(u-\frac{\lambda}{2}i\right)x} h_K(x-2\pi)\,\mathrm{d}x\right)\right|^2 \mathrm{d}u \\ + \quad & c\int \left|\frac{3}{4} + \frac{u^2}{\lambda^2}\right|^{-\beta} \left|1 - \exp\left(\gamma \int e^{i\left(u+\frac{\lambda}{2}i\right)x} h_K(x-2\pi)\,\mathrm{d}x\right)\right|^2 \mathrm{d}u \\ = \quad & c\int \left|\frac{3}{4} + \frac{u^2}{\lambda^2}\right|^{-\beta} \left|1 - \exp\left(\gamma e^{i(u-\lambda i/2)2\pi} \mathcal{F}h_K\left(u - \frac{\lambda i}{2}\right)\right)\right|^2 \mathrm{d}u \\ + \quad & c\int \left|\frac{3}{4} + \frac{u^2}{\lambda^2}\right|^{-\beta} \left|1 - \exp\left(\gamma e^{i(u+\lambda i/2)2\pi} \mathcal{F}h_K\left(u + \frac{\lambda i}{2}\right)\right)\right|^2 \mathrm{d}u. \end{aligned} \tag{2.4.31}$$

Finally, using the fact that $|e^z - 1| \leq |z|e^{|\operatorname{Re}(z)|}$ for arbitrary $z \in \mathbb{C}$ and that h_K is supported in $\left[\frac{3\pi}{2}, \frac{5\pi}{2}\right]$, we find that

$$\begin{aligned} & c\int \left|\frac{3}{4} + \frac{u^2}{\lambda^2}\right|^{-\beta} \left|1 - \exp\left(\gamma e^{i(u-\lambda i/2)2\pi} \mathcal{F}h_K\left(u - \frac{\lambda i}{2}\right)\right)\right|^2 \mathrm{d}u \\ & \leq c\exp\left(2\gamma e^{3\lambda\pi}\|h_k\|_{\mathrm{L}^1}\right)\gamma e^{\lambda\pi} \int \left|\frac{3}{4} + \frac{u^2}{\lambda^2}\right|^{-\beta} \left|\mathcal{F}h_K\left(u - \frac{\lambda i}{2}\right)\right|^2 \mathrm{d}u \\ & = c\exp\left(2\gamma e^{3\lambda\pi}\|h_k\|_{\mathrm{L}^1}\right)\gamma e^{\lambda\pi} \\ & \qquad \int K^{-2a-2} \left|\frac{3}{4} + \frac{u^2}{\lambda^2}\right|^{-\beta} \left|\mathcal{F}h_0\left(\frac{u + \lambda i/2}{K}\right)\right|^2 \mathrm{d}u \\ & \asymp K^{-2a-2\beta-1}. \end{aligned} \tag{2.4.32}$$

The same arguments are used to see that

$$\begin{aligned} & c\int \left|\frac{3}{4} + \frac{u^2}{\lambda^2}\right|^{-\beta} \left|1 - \exp\left(\gamma e^{i(u-\lambda i/2)2\pi} \mathcal{F}h_K\left(u - \frac{\lambda i}{2}\right)\right)\right|^2 \mathrm{d}u \\ & \asymp K^{-2a-2\beta-1}. \end{aligned} \tag{2.4.33}$$

We have thus shown that for any sequence $K_n \to \infty$, the distributions $\mathbb{P}_0$ and

$\mathbb{P}_{K_n}$ do not separate, provided that $K_n \asymp n^{\frac{1}{2a+2\beta+1}}$.

Now, Lemma 2.4.3 and Theorem 2.4.5 imply that for $r_n = n^{\frac{2a+2}{2\beta+2a+1}}$,

$$\liminf_{n\to\infty} \inf_{\bar{\theta}_n} \sup_{\mu\in\mathcal{M}} \mathbb{E}_\mu \left[\left| \theta - \bar{\theta}_n \right|^2 \right] r_n \geq c, \tag{2.4.34}$$

which is the desired result. □

In a similar manner, we obtain the result for estimating the Lévy density with pointwise loss:

Theorem 2.4.7. *Let $\mathcal{M}$ be defined as in the preceding theorem. Let $y \in \mathbb{R}\setminus\{0\}$ and let $f(x) = y^{-1}\delta_y(x)$. Then we have for $\theta = \int f(x)\mu(\,dx) = \eta(y)$:*

$$\liminf_{n\to\infty} \inf_{\bar{\theta}_n} \sup_{\mu\in\mathcal{M}} \mathbb{E}_\mu \left[\left| \theta - \bar{\theta}_{\Delta,n} \right|^2 \right] T^{\frac{2a}{2a+2\Delta\beta+1}} > 0. \tag{2.4.35}$$

Proof. Without loss of generality, let $y = 2\pi$ and $D = (\pi, 3\pi)$. We consider, again, the same bilateral Gamma distribution $\mathbb{P}_0$ as in the proof of the preceding statement. Again, let h_0 be antisymmetric around the origin with support in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and assume, this time, that $h_0(\frac{\pi}{4}) > 0$. Let $h_K(x) = K^{-a}h_0(Kx)$ and set, this time

$$g_K(x) = g(x) + \gamma x h_K\left(\left(x - \left(2\pi - \frac{\pi}{4K}\right)\right)\right). \tag{2.4.36}$$

Again, we may chose h_0 Hölder regular and argue along the same lines as in the proof of the preceding statement that the signed measure μ_K with density g_K belongs to $\mathcal{M}$. The loss, can be calculated, this time, as follows:

$$\begin{aligned} y^{-1}\left|g_0(y) - g_K(y)\right| &= \left|\eta_0(y) - \eta_K(y)\right| \\ &= \left|\gamma h_K\left(\frac{\pi}{4K}\right)\right| = \gamma K^{-a} h_0\left(\frac{\pi}{4}\right). \end{aligned} \tag{2.4.37}$$

Calculating the χ^2-distance runs exactly along the same lines as in the proof of the preceding statement. From this we derive that

$$\limsup_{n\to\infty} \inf_{\bar{\theta}_n} \sup_{\mu\in\mathcal{M}} \mathbb{E}_\mu \left[|\theta - \bar{\theta}_n|^2 \right] n^{\frac{2a}{2a+2\beta+1}} > 0, \tag{2.4.38}$$

and hence the statement of the theorem. □

2.5 Proofs

2.5.1 Proofs of the main results of Section 2.2

We start by proving the following key Lemma:

Lemma 2.5.1.

a) Given discrete, equidistant observations of a pure jump Lévy process satisfying the assumptions (A1)-(A4), let $\widehat{\varphi}'_{\Delta,n}$ and $\frac{1}{\widetilde{\varphi}_{\Delta,n}}$ be defined by (2.2.14) and (2.2.15). Then we can estimate

$$\left|\mathbb{E}\left[\left(\frac{\frac{1}{\Delta}\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\frac{1}{\Delta}\varphi'_{\Delta}(u)}{\varphi_{\Delta}(u)}\right)\left(\frac{\frac{1}{\Delta}\widehat{\varphi}'_{\Delta,n}(-v)}{\widetilde{\varphi}_{\Delta,n}(-v)}-\frac{\frac{1}{\Delta}\varphi'_{\Delta}(-v)}{\varphi_{\Delta}(-v)}\right)\right]\right|$$
$$\leq C\left(\frac{T^{-1}}{|\varphi_{\Delta}(u)\varphi_{\Delta}(-v)|}\wedge 1\right)\Big(|\Psi''(u-v)|+|\Psi'(u-v)|^2+|\Psi'(u)\Psi'(-v)|\Big),$$

with $C := 1+2C_1+2C_2$, where C_1 and C_2 denote the constants in Lemma 2.2.2.

b) In the density deconvolution model, let $\widehat{\varphi}_{Z_n}$ and $\frac{1}{\widetilde{\varphi}_{\varepsilon n}} =: \frac{1}{\widetilde{\varphi}_n}$ be defined by (2.2.4) and (2.2.6). Then we can estimate

$$\left|\mathbb{E}\left(\frac{\widehat{\varphi}_{Z_n}(u)}{\widetilde{\varphi}_n(u)}-\frac{\varphi_Z(u)}{\varphi(u)}\right)\left(\frac{\widehat{\varphi}_{Z_n}(-v)}{\widetilde{\varphi}_n(-v)}-\frac{\varphi_Z(-v)}{\varphi(-v)}\right)\right|$$
$$\leq C\left(\frac{n^{-1}}{|\varphi(u)\varphi(-v)|}\wedge 1\right)\Big(|\varphi_Y(u-v)|+|\varphi_Y(u-v)|^2+|\varphi_Y(u)\varphi_Y(-v)|\Big).$$

Proof. We give a rigorous proof of the a) part and then sketch the proof of part b).

We start by noticing that we have

$$\mathbb{E}\left[\left|\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{1}{\varphi_{\Delta}(u)}\right|^k\right]\leq C_k\left(\frac{T^{-\frac{k}{2}}}{|\varphi_{\Delta}(u)|^{2k}}\wedge\frac{1}{|\varphi_{\Delta}(u)|^k}\right). \tag{2.5.1}$$

This is a direct consequence of Lemma 2.2.2, tracing back the dependence on Δ.

We can write

$$\mathbb{E}\left[\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\varphi'_{\Delta}(u)}{\varphi_{\Delta}(u)}\right)\left(\frac{\widehat{\varphi}'_{\Delta,n}(-v)}{\widetilde{\varphi}_{\Delta,n}(-v)}-\frac{\varphi'_{\Delta}(-v)}{\varphi_{\Delta}(-v)}\right)\right]$$
$$=\mathbb{E}\Bigg[\left(\frac{(\widehat{\varphi}'_{\Delta,n}(u)-\varphi'_{\Delta}(u))}{\widetilde{\varphi}_{\Delta,n}(u)}+\varphi'_{\Delta}(u)\left(\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{1}{\varphi_{\Delta}(u)}\right)\right)$$
$$\left(\frac{(\widehat{\varphi}'_{\Delta,n}(-v)-\varphi'_{\Delta}(-v))}{\widetilde{\varphi}_{\Delta,n}(-v)}+\varphi'_{\Delta}(-v)\left(\frac{1}{\widetilde{\varphi}_{\Delta,n}(-v)}-\frac{1}{\varphi_{\Delta}(-v)}\right)\right)\Bigg]. \tag{2.5.2}$$

Using the fact that $\widehat{\varphi}'_{\Delta,n}$ and $\frac{1}{\widetilde{\varphi}_{\Delta,n}}$ are independent by construction and that

$\widehat{\varphi}'_{\Delta,n}(u) - \varphi'_\Delta(u)$ and $\widehat{\varphi}'_{\Delta,n}(-v) - \varphi'_\Delta(-v)$ are centred, we find that

$$\begin{aligned}
&\mathbb{E}\Bigg[\left(\frac{(\widehat{\varphi}'_{\Delta,n}(u) - \varphi'_\Delta(u))}{\widetilde{\varphi}_{\Delta,n}(u)} + \varphi'_\Delta(u)\left(\frac{1}{\widetilde{\varphi}_{\Delta,n(u)}} - \frac{1}{\varphi_\Delta(u)}\right)\right) \\
&\qquad \left(\frac{(\widehat{\varphi}'_{\Delta,n}(-v) - \varphi'_\Delta(-v))}{\widetilde{\varphi}_{\Delta,n}(-v)} + \varphi'_\Delta(-v)\left(\frac{1}{\widetilde{\varphi}_{\Delta,n}(-v)} - \frac{1}{\varphi_\Delta(-v)}\right)\right)\Bigg] \\
= \;& \mathbb{E}\Big[\Big(\widehat{\varphi}'_{\Delta,n}(u) - \varphi'_\Delta(u)\Big)\Big(\widehat{\varphi}'_{\Delta,n}(-v) - \varphi'_\Delta(-v)\Big)\Big]\mathbb{E}\left[\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)\widetilde{\varphi}_{\Delta,n}(-v)}\right] \\
+ \;& \varphi'_\Delta(u)\varphi'_\Delta(-v)\,\mathbb{E}\left[\left(\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)} - \frac{1}{\varphi_\Delta(u)}\right)\left(\frac{1}{\widetilde{\varphi}_{\Delta,n}(-v)} - \frac{1}{\varphi_\Delta(-v)}\right)\right] \\
= \;& \mathrm{Cov}\left(\widehat{\varphi}'_{\Delta,n}(u), \widehat{\varphi}'_{\Delta,n}(-v)\right)\mathbb{E}\left[\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)\widetilde{\varphi}_{\Delta,n}(-v)}\right] \\
+ \;& \varphi'_\Delta(u)\varphi'_\Delta(-v)\,\mathbb{E}\left[\left(\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)} - \frac{1}{\varphi_\Delta(u)}\right)\left(\frac{1}{\widetilde{\varphi}_{\Delta,n}(-v)} - \frac{1}{\varphi_\Delta(-v)}\right)\right] \\
=: \;& \mathrm{Cov}\left(\widehat{\varphi}'_{\Delta,n}(u), \widehat{\varphi}'_{\Delta,n}(-v)\right)\mathbb{E}\left[\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)\widetilde{\varphi}_{\Delta,n}(-v)}\right] \\
+ \;& \varphi'_\Delta(u)\varphi'_\Delta(-v)\,\mathbb{E}\left[\mathrm{R}_{\Delta,n}(u)\mathrm{R}_{\Delta,n}(-v)\right].
\end{aligned} \tag{2.5.3}$$

The Cauchy-Schwarz-inequality and then an application of (2.5.1) imply

$$\begin{aligned}
&\mathbb{E}\left[|\mathrm{R}_{\Delta,n}(u)\mathrm{R}_{\Delta,n}(-v)|\right] \\
\leq \;& \left(\mathbb{E}\left[|\mathrm{R}_{\Delta,n}(u)|^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[|\mathrm{R}_{\Delta,n}(-v)|^2\right]\right)^{\frac{1}{2}} \\
\leq \;& C_2\left(\frac{T^{-1}}{|\varphi_\Delta(u)|^2|\varphi_\Delta(-v)|^2} \wedge \frac{1}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|}\right).
\end{aligned} \tag{2.5.4}$$

Next, using the triangle inequality, again (2.5.1) and then (2.5.4), we find that

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)\widetilde{\varphi}_{\Delta,n}(-v)}\right|\right] \\
\leq \;& \mathbb{E}\left[\left(|\mathrm{R}_{\Delta,n}(u)| + \frac{1}{|\varphi_\Delta(u)|}\right)\left(|\mathrm{R}_{\Delta,n}(-v)| + \frac{1}{|\varphi_\Delta(-v)|}\right)\right] \\
\leq \;& \frac{1}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|} + \frac{1}{|\varphi_\Delta(u)|}\mathbb{E}\left[|\mathrm{R}_{\Delta,n}(-v)\right] \\
+ \;& \frac{1}{|\varphi_\Delta(-v)|}\mathbb{E}\left[|\mathrm{R}_{\Delta,n}(u)|\right] + \mathbb{E}\left[|\mathrm{R}_{\Delta,n}(u)|\,|\mathrm{R}_{\Delta,n}(-v)|\right] \\
\leq \;& (1 + 2C_1 + C_2)\frac{1}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|}.
\end{aligned} \tag{2.5.5}$$

Moreover, by definition of $\frac{1}{\widetilde{\varphi}_n}$, we have

$$\mathbb{E}\left[\left|\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)\widetilde{\varphi}_{\Delta,n}(-v)}\right|\right] \leq (\Delta n) = T. \tag{2.5.6}$$

Next, we calculate

$$\begin{aligned}
& \operatorname{Cov}(\widehat{\varphi}'_{\Delta,n}(u), \widehat{\varphi}'_{\Delta,n}(v)) \\
= \;& n^{-1}\left(\mathbb{E}\left[(iZ_\Delta)^2 e^{i(u-v)Z_\Delta}\right] - \mathbb{E}\left[iZ_\Delta e^{iuZ_\Delta}\right]\mathbb{E}\left[iZ_\Delta e^{-ivZ_\Delta}\right]\right) \\
= \;& n^{-1}\left(\varphi''_\Delta(u-v) - \varphi'_\Delta(u)\varphi'_\Delta(-v)\right).
\end{aligned} \tag{2.5.7}$$

Moreover, we clearly have

$$\begin{aligned}
|\varphi'_\Delta(u)|\,|\varphi'_\Delta(-v)| &= |\Delta\Psi'(u)\varphi_\Delta(u)|\,|\Delta\Psi'(-v)\varphi_\Delta(-v)| \\
&\leq |\Delta\Psi'(u)||\Delta\Psi'(-v)|,
\end{aligned} \tag{2.5.8}$$

and

$$\begin{aligned}
|\varphi''_\Delta(u-v)| &= \left|\Delta\Psi''(u-v)\varphi_\Delta(u-v) + \Delta^2(\Psi'(u-v))^2\varphi_\Delta(u-v)\right| \\
&\leq \Delta|\Psi''(u-v)| + \Delta^2(\Psi'(u-v))^2.
\end{aligned} \tag{2.5.9}$$

Putting (2.5.7)-(2.5.9) and (2.5.5) together, we have shown that

$$\begin{aligned}
& \left|\operatorname{Cov}\left(\widehat{\varphi}'_{\Delta,n}(u), \widehat{\varphi}'_{\Delta,n}(v)\right)\right|\left|\mathbb{E}\left[\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)\widetilde{\varphi}_{\Delta,n}(-v)}\right]\right| \\
& \leq C'\frac{n^{-1}\Delta}{|\varphi_\Delta(u)\varphi_\Delta(-v)|}\left(|\Psi''(u-v)| + \Delta|\Psi'(u-v)|^2 + |\Delta\Psi'(u)\Psi'(-v)|\right) \\
& \leq C'\frac{\Delta^2 T^{-1}}{|\varphi_\Delta(u)\varphi_\Delta(-v)|}\left(|\Psi''(u-v)| + |\Psi'(u-v)|^2 + |\Psi'(u)\Psi'(-v)|\right).
\end{aligned} \tag{2.5.10}$$

With constant $C' := 1 + 2C_2 + C_1$.

On the other hand, formulae (2.5.7)-(2.5.9) and (2.5.6) imply

$$\begin{aligned}
& \left|\operatorname{Cov}\left(\widehat{\varphi}'_{\Delta,n}(u), \widehat{\varphi}'_{\Delta,n}(v)\right)\right|\left|\mathbb{E}\left[\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)\widetilde{\varphi}_{\Delta,n}(-v)}\right]\right| \\
& \leq C'Tn^{-1}\Big(|\Delta\Psi''(u-v)| + |\Delta\Psi'(u-v)|^2 + \left|\Delta^2\Psi'(u)\Psi'(-v)\right|\Big) \\
& \leq C'\Delta^2\Big(|\Psi''(u-v)| + |\Psi'(u-v)|^2 + |\Psi'(u)\Psi'(-v)|\Big).
\end{aligned} \tag{2.5.11}$$

We have thus shown that the following estimate holds:

$$\begin{aligned}
& \left|\operatorname{Cov}\left(\widehat{\varphi}'_{\Delta,n}(u), \widehat{\varphi}'_{\Delta,n}(v)\right)\right|\left|\mathbb{E}\left[\frac{1}{\widetilde{\varphi}_{\Delta,n}(u)\widetilde{\varphi}_{\Delta,n}(-v)}\right]\right| \\
& \leq \Delta^2 C'\left(1 \wedge \frac{T^{-1}}{|\varphi_\Delta(u)\varphi_\Delta(-v)|}\right)
\end{aligned}$$

$$\left(\left|\Psi''(u-v)\right| + \left|\Psi'(u-v)\right|^2 + \left|\Psi'(u)\Psi'(-v)\right|\right). \quad (2.5.12)$$

It remains to consider the last line of formula (2.5.3). Using (2.5.4) and then the fact that $\varphi'_\Delta(u) = \Delta\Psi'(u)\varphi_\Delta(u)$, we can estimate

$$\begin{aligned}
&|\varphi'_\Delta(u)||\varphi'_\Delta(-v)| \left|\mathbb{E}\left[\mathrm{R}_{\Delta,n}(u)\mathrm{R}_{\Delta,n}(-v)\right]\right| \\
&\leq C_2 \left(\frac{T^{-1}}{|\varphi_\Delta(u)|^2|\varphi_\Delta(-v)|^2} \wedge \frac{1}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|}\right)|\varphi'_\Delta(u)||\varphi'_\Delta(-v)| \\
&= C_2 \left(\frac{T^{-1}}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|} \wedge 1\right) \Delta^2 \left|\Psi'(u)\right| \left|\Psi'(-v)\right|. \qquad (2.5.13)
\end{aligned}$$

Putting (2.5.2) and (2.5.3), (2.5.12) and (2.5.13) together, we have shown that

$$\begin{aligned}
&\frac{1}{\Delta^2}\left|\mathbb{E}\left[\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)} - \frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)}\right)\left(\frac{\widehat{\varphi}'_{\Delta,n}(-v)}{\widetilde{\varphi}_{\Delta,n}(-v)} - \frac{\varphi'_\Delta(-v)}{\varphi_\Delta(-v)}\right)\right]\right| \\
&\leq C\left(\frac{T^{-1}}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|} \wedge 1\right)\left(\left|\Psi''(u-v)\right| + |\Psi'(u-v)|^2 + |\Psi'(u)||\Psi'(-v)|\right), \\
&\qquad (2.5.14)
\end{aligned}$$

which is the statement of the a) part.

To see the b) part, we can argue exactly along the same lines, replacing, in each step, $\widehat{\varphi}'_{\Delta,n}$ by $\widehat{\varphi}_{Z_n}$ and $\frac{1}{\widetilde{\varphi}_{\Delta,n}}$ by $\frac{1}{\widetilde{\varphi}_{\varepsilon n}}$.

In formula (2.5.7), we have to use the fact that

$$\begin{aligned}
&\mathrm{Cov}\left(\widehat{\varphi}_{Z_1}(u), \widehat{\varphi}_{Z_1}(v)\right) \\
&= \mathbb{E}\left[e^{i(u-v)Z_1}\right] - \mathbb{E}\left[e^{iuZ_1}\right]\mathbb{E}\left[e^{-ivZ_1}\right] \\
&= \varphi_Z(u-v) - \varphi_Z(u)\varphi_Z(-v) \qquad (2.5.15)
\end{aligned}$$

and in (2.5.13), we have to use the fact that $\varphi_Z = \varphi_Y\varphi_\varepsilon$ and hence $\frac{\varphi_Z}{\varphi_\varepsilon} = \varphi_Y$. Finally, we trivially have $|\varphi_Z| = |\varphi_\varepsilon\varphi_Y| \leq |\varphi_Y|$. This reasoning gives the statement of part b). $\square$

We are now ready to prove Theorem 2.2.9 and Theorem 2.2.10.

Proof of Theorem 2.2.9. The risk of $\widehat{\theta}_{\Delta,h,n}$ can be decomposed as follows: Given the kernel function K and bandwidth h, let $\theta_h := \int f(x)(\mathrm{K}_h * \mu)(x)\,\mathrm{d}x$. We can estimate

$$\begin{aligned}
&\mathbb{E}\left[\left|\theta - \widehat{\theta}_{\Delta,h,n}\right|^2\right] \\
&\leq 2\left|\theta - \theta_h\right|^2 + 2\,\mathbb{E}\left[\left|\theta_h - \widehat{\theta}_{\Delta,h,n}\right|^2\right] \\
&= 2\left|\int f(x)\mu(\,\mathrm{d}x) - \int f(x)(\mathrm{K}_h * \mu)(x)\,\mathrm{d}x\right|^2 \\
&+2\,\mathbb{E}\left[\left|\int f(x)(\mathrm{K}_h * \mu)(x)\,\mathrm{d}x - \frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\,\mathrm{K}_h(u)\frac{\frac{1}{\Delta}\widehat{\varphi}'_{\Delta,n}(u)}{i\widetilde{\varphi}_{\Delta,n}(u)}\,\mathrm{d}u\right|^2\right]. (2.5.16)
\end{aligned}$$

By assumption on K, we have $\mathcal{F}K_h\mathcal{F}f \in \mathrm{L}^1(\mathbb{R})$, so we can pass to the Fourier domain and find that

$$\begin{aligned}
&\mathbb{E}\left[\left|\int f(x)(\mathrm{K}_h * \mu)(x)\,\mathrm{d}x - \frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\,\mathrm{K}_h(u)\frac{\frac{1}{\Delta}\widehat{\varphi}_{\Delta,n}(u)}{i\widetilde{\varphi}_{\Delta,n}(u)}\,\mathrm{d}u\right|^2\right] \\
&= \mathbb{E}\left[\left|\frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\,\mathrm{K}_h(u)\mathcal{F}\mu(u)\,\mathrm{d}u - \int \mathcal{F}f(-u)\mathcal{F}\,\mathrm{K}_h(u)\frac{\frac{1}{\Delta}\widehat{\varphi}_{\Delta,n}(u)}{i\widetilde{\varphi}_{\Delta,n}(u)}\,\mathrm{d}u\right|^2\right] \\
&= \mathbb{E}\left[\left|\frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\,\mathrm{K}_h(u)\frac{1}{\Delta}\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)} - \frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)}\right)\mathrm{d}u\right|^2\right]. \qquad (2.5.17)
\end{aligned}$$

Using the fact that X_Δ has, by assumption, a finite second moment, that the absolute value of $\frac{1}{\widetilde{\varphi}_n}$ is bounded above and that $\varphi'(u)/\varphi(u) = \Psi'(u)$ is uniformly bounded in u, and once again the fact that $\mathcal{F}\,\mathrm{K}(-\bullet)\mathcal{F}f \in \mathrm{L}^1(\mathbb{R})$, we find that Fubini's theorem applies. We can thus estimate

$$\begin{aligned}
&\mathbb{E}\left[\left|\int \mathcal{F}f(-u)\mathcal{F}\,\mathrm{K}_h(u)\frac{1}{\Delta}\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)} - \frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)}\right)\mathrm{d}u\right|^2\right] \\
&= \mathbb{E}\Bigg[\int\int \mathcal{F}f(-u)\mathcal{F}f(v)\mathcal{F}\,\mathrm{K}_h(u)\mathcal{F}\,\mathrm{K}_h(-v) \\
&\qquad \frac{1}{\Delta^2}\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)} - \frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)}\right)\left(\frac{\widehat{\varphi}'_{\Delta,n}(-v)}{\widetilde{\varphi}_{\Delta,n}(-v)} - \frac{\varphi'_\Delta(-v)}{\varphi_\Delta(-v)}\right)\mathrm{d}u\,\mathrm{d}v\Bigg] \\
&= \int\int \mathcal{F}f(-u)\mathcal{F}f(v)\mathcal{F}\,\mathrm{K}_h(u)\mathcal{F}\,\mathrm{K}_h(-v) \\
&\qquad \frac{1}{\Delta^2}\,\mathbb{E}\left[\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)} - \frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)}\right)\left(\frac{\widehat{\varphi}'_{\Delta,n}(-v)}{\widetilde{\varphi}_{\Delta,n}(-v)} - \frac{\varphi'_\Delta(-v)}{\varphi_\Delta(-v)}\right)\right]\mathrm{d}u\,\mathrm{d}v. \qquad (2.5.18)
\end{aligned}$$

Next, Lemma 2.5.1 gives

$$\begin{aligned}
&\iint \mathcal{F}f(-u)\mathcal{F}f(v)\mathcal{F}\,\mathrm{K}_h(u)\mathcal{F}\,\mathrm{K}_h(-v) \\
&\qquad \frac{1}{\Delta^2}\,\mathbb{E}\left[\left(\frac{\widehat{\varphi}_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)} - \frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)}\right)\left(\frac{\widehat{\varphi}_{\Delta,n}(-v)}{\widetilde{\varphi}_{\Delta,n}(-v)} - \frac{\varphi'_\Delta(-v)}{\varphi_\Delta(-v)}\right)\right]\mathrm{d}u\,\mathrm{d}v \\
&\le CT^{-1}\Bigg\{\iint \frac{|\mathcal{F}f(-u)\mathcal{F}f(v)|}{|\varphi_\Delta(u)\varphi_\Delta(-v)|}\left|\Psi'(u)\Psi'(-v)\right|\mathrm{d}u\,\mathrm{d}v \\
&+\iint \frac{|\mathcal{F}f(-u)\mathcal{F}f(v)|}{|\varphi_\Delta(u)\varphi_\Delta(-v)|}|\mathrm{K}_h(u)\,\mathrm{K}_h(-v)|\Big(\left|\Psi''(u-v)\right| + \left|\Psi'(u-v)\right|^2\Big)\,\mathrm{d}u\,\mathrm{d}v\Bigg\}. (2.5.19)
\end{aligned}$$

If $\Psi'' \in \mathrm{L}^1(\mathbb{R})$ and $\Psi' \in \mathrm{L}^2(\mathbb{R})$ hold true, we apply the Cauchy-Schwarz inequality and Fubini's theorem to find that

$$\int\int \frac{|\mathcal{F}f(-u)|\,|\mathcal{F}f(v)|}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|}\,|\mathrm{K}_h(u)|\,|\mathrm{K}_h(-v)|\left(\left|\Psi''(u-v)\right| + \left|\Psi'(u-v)\right|^2\right)\mathrm{d}u\,\mathrm{d}v$$

$$
\begin{aligned}
&\le \int\int \frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2} |\mathcal{F}\,\mathrm{K}_h(u)|^2 \left(|\Psi''(u-v)| + |\Psi'(u-v)|^2\right) \mathrm{d}u\,\mathrm{d}v \\
&= \int \frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2} |\mathcal{F}\,\mathrm{K}_h(u)|^2 \int \left(|\Psi''(u-v)| + |\Psi'(u-v)|^2\right) \mathrm{d}v\,\mathrm{d}u \\
&\le \sup_{u\in\mathbb{R}} \int \left(|\Psi''_\Delta(u-v)| + |\Psi'(u-v)|^2\right) \mathrm{d}v \int \frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2} |\mathcal{F}\,\mathrm{K}_h(u)|^2 \,\mathrm{d}u \\
&\le \left(\int (|\Psi''(x)|\,\mathrm{d}x + \int |\Psi'(x)|^2\,\mathrm{d}x\right) \int \frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2} |\mathcal{F}\,\mathrm{K}_h(u)|^2 \,\mathrm{d}u. \qquad (2.5.20)
\end{aligned}
$$

Another application of the Cauchy-Schwarz-inequality gives

$$
\begin{aligned}
&\int\int \frac{|\mathcal{F}f(-u)||\mathcal{F}f(v)|}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|} |\mathrm{K}_h(u)|\,|\mathrm{K}_h(-v)|\,|\Psi'(u)||\Psi'(-v)|\,\mathrm{d}u\,\mathrm{d}v \\
&= \left(\int \frac{|\mathcal{F}f(-u)|}{|\varphi_\Delta(u)|} |\mathcal{F}\,\mathrm{K}_h(u)||\Psi'(u)|\,\mathrm{d}u\right)^2 \\
&\le \int |\Psi'(u)|^2\,\mathrm{d}u \int \frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2} |\mathcal{F}\,\mathrm{K}_h(u)|^2\,\mathrm{d}u. \qquad (2.5.21)
\end{aligned}
$$

We have thus shown that

$$
\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{2\pi}\int f(x)(\mathrm{K}_h * \mu)(x)\,\mathrm{d}x - \frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\,\mathrm{K}_h(u)\frac{\frac{1}{\Delta}\widehat{\varphi}_{\Delta,n}(u)}{i\widetilde{\varphi}_{\Delta,n}(u)}\,\mathrm{d}u\right|^2\right] \\
&\le \frac{C}{(2\pi)^2} T^{-1}\left(\int |\Psi''(x)|\,\mathrm{d}x + 2\int |\Psi'(x)|^2\,\mathrm{d}x\right)\int \frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2} |\mathcal{F}\,\mathrm{K}_h(u)|^2\,\mathrm{d}u. \qquad (2.5.22)
\end{aligned}
$$

On the other hand, if $\Psi'' \in \mathrm{L}^1(\mathbb{R})$ or $\Psi' \in \mathrm{L}^2(\mathbb{R})$ fails to hold, we can always estimate

$$
\begin{aligned}
&\iint \frac{|\mathcal{F}f(-u)|\,|\mathcal{F}f(v)|}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|} |\mathrm{K}_h(u)|\,|\mathrm{K}_h(-v)| \left(|\Psi''(u-v)| + |\Psi'(u-v)|^2\right) \mathrm{d}u\,\mathrm{d}v \\
&\le \sup_{u,v\in\mathbb{R}} \left(|\Psi''(u-v)| + |\Psi'(u-v)|^2\right) \iint \frac{|\mathcal{F}f(-u)|\,|\mathcal{F}f(v)|}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|} |\mathrm{K}_h(u)|\,|\mathrm{K}_h(-v)|\,\mathrm{d}u\,\mathrm{d}v \\
&\le \sup_{x\in\mathbb{R}} \left(|\Psi''(x)| + |\Psi'(x)|^2\right) \left(\int \frac{|\mathcal{F}f(-u)|}{|\varphi_\Delta(u)|} |\mathcal{F}\,\mathrm{K}_h(u)|\,\mathrm{d}u\right)^2 \qquad (2.5.23)
\end{aligned}
$$

and

$$
\begin{aligned}
&\int\int \frac{|\mathcal{F}f(-u)||\mathcal{F}f(v)|}{|\varphi_\Delta(u)||\varphi_\Delta(-v)|} |\mathcal{F}\,\mathrm{K}_h(u)||\mathcal{F}\,\mathrm{K}_h(-v)||\Psi'(u)||\Psi'(-v)|\,\mathrm{d}u\,\mathrm{d}v \\
&= \left(\int \frac{|\mathcal{F}f(-u)|}{|\varphi_\Delta(u)|} |\mathcal{F}\,\mathrm{K}_h(u)||\Psi'(u)|\,\mathrm{d}u\right)^2 \\
&\le \sup_{x\in\mathbb{R}} |\Psi'(x)|^2 \left(\int \frac{|\mathcal{F}f(-u)|}{|\varphi_\Delta(u)|} |\mathcal{F}\,\mathrm{K}_h(u)|\,\mathrm{d}u\right)^2. \qquad (2.5.24)
\end{aligned}
$$

This yields

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{2\pi}\int f(x)(\mathrm{K}_h*\mu)(x)\,\mathrm{d}x-\frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\mathrm{K}_h(u)\frac{\frac{1}{\Delta}\widehat{\varphi}_{\Delta_n}(u)}{i\widetilde{\varphi}_{\Delta_n}(u)}\,\mathrm{d}u\right|^2\right]\\
&\leq\frac{C}{(2\pi)^2}T^{-1}\left(\sup_{x\in\mathbb{R}}|\Psi''(x)|+2\sup_{x\in\mathbb{R}}|\Psi'(x)|^2\right)\left(\int\frac{|\mathcal{F}f(-u)|}{|\varphi_\Delta(u)|}|\mathcal{F}\mathrm{K}_h(u)|\,\mathrm{d}u\right)^2.
\end{aligned} \tag{2.5.25}$$

Putting the above results together, we have shown that

$$\begin{aligned}
&\mathbb{E}\left[\left|\theta-\widehat{\theta}_{\Delta,h,n}\right|^2\right]\\
&\leq 2\left|\int f(x)\mu(\,\mathrm{d}x)-\int f(x)(\mathrm{K}_h*\mu)(x)\,\mathrm{d}x\right|^2\\
&+\frac{C}{2\pi^2}T^{-1}\Bigg\{\left(\|\Psi''(x)\|_{\mathrm{L}^1}\,\mathrm{d}x+2\|\Psi'(x)\|_{\mathrm{L}^2}^2\right)\int\frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2}|\mathcal{F}\mathrm{K}_h(u)|^2\,\mathrm{d}u\\
&\qquad\wedge\left(\|\Psi''(x)\|_\infty+2\|\Psi'(x)\|_\infty^2\right)\left(\int\frac{|\mathcal{F}f(-u)|}{|\varphi_\Delta(u)|}|\mathcal{F}\mathrm{K}_h(u)|\,\mathrm{d}u\right)^2\Bigg\}.
\end{aligned} \tag{2.5.26}$$

This is the statement of the theorem. ☐

Next, we prove Theorem 2.2.10.

Proof of Theorem 2.2.10. First, recall that we now assume that $|\mathcal{F}f|\in\mathrm{L}^1(\mathbb{R})$, so we certainly have $|\mathcal{F}f\mathcal{F}\mu|\leq\|\mathcal{F}\mu\|_\infty|\mathcal{F}f|\in\mathrm{L}^1(\mathbb{R})$. We can thus express θ in the Fourier domain, writing

$$\theta=\int f(x)\mu(\,\mathrm{d}x)=\frac{1}{2\pi}\int\mathcal{F}f(-u)\mathcal{F}\mu(u)\,\mathrm{d}u. \tag{2.5.27}$$

The squared risk of $\widehat{\theta}_{\Delta,n}$ can be estimated as follows:

$$\begin{aligned}
&\mathbb{E}\left[\left|\theta-\widehat{\theta}_n\right|^2\right]\\
=\;&\mathbb{E}\left[\left|\int f(x)\mu(\,\mathrm{d}x)-\frac{1}{2\pi}\int\mathcal{F}f(-u)\frac{\frac{1}{\Delta}\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}\,\mathrm{d}u\right|^2\right]\\
=\;&\mathbb{E}\left[\left|\frac{1}{2\pi}\int\mathcal{F}f(-u)\mathcal{F}\mu(u)\,\mathrm{d}u-\frac{1}{2\pi}\int\mathcal{F}f(-u)\frac{\frac{1}{\Delta}\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}\,\mathrm{d}u\right|^2\right]\\
=\;&\mathbb{E}\left[\left|\frac{1}{2\pi}\int\mathcal{F}f(-u)\frac{1}{\Delta}\left(\frac{\widehat{\varphi}_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)}\right)\,\mathrm{d}u\right|^2\right].
\end{aligned} \tag{2.5.28}$$

For arbitrary $m\geq 0$, we can estimate

$$\mathbb{E}\left[\left|\frac{1}{2\pi}\int\mathcal{F}f(-u)\frac{1}{\Delta}\left(\frac{\widehat{\varphi}_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)}\right)\,\mathrm{d}u\right|^2\right]$$

$$\leq \; 2\,\mathbb{E}\left[\left|\frac{1}{2\pi}\int\limits_{\{|u|>\pi m\}} \mathcal{F}f(-u)\frac{1}{\Delta}\left(\frac{\widehat{\varphi}_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\varphi'_{\Delta}(u)}{\varphi_{\Delta}(u)}\right)\mathrm{d}u\right|^2\right]$$
$$+ \; 2\,\mathbb{E}\left[\left|\frac{1}{2\pi}\int\limits_{\{|u|\leq\pi m\}} \mathcal{F}f(-u)\frac{1}{\Delta}\left(\frac{\widehat{\varphi}_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\varphi'_{\Delta}(u)}{\varphi_{\Delta}(u)}\right)\mathrm{d}u\right|^2\right]. \quad (2.5.29)$$

The expression appearing in the last line of formula (2.5.29) is a special case of the expression appearing in the last line of (2.5.16), with K_h denoting the sinc kernel. We can thus argue exactly along the same lines as in formulae (2.5.20)-(2.5.24) and derive from repeated applications of the Cauchy-Schwarz inequality and Fubini's theorem and from Lemma 2.5.1 that

$$\mathbb{E}\left[\left|\frac{1}{2\pi}\int\limits_{\{|u|\leq\pi m\}} \mathcal{F}f(-u)\frac{1}{\Delta}\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\varphi'_{\Delta}(u)}{\varphi_{\Delta}(u)}\right)\mathrm{d}u\right|^2\right]$$
$$= \frac{1}{(2\pi)^2}\int\limits_{\{|u|\leq\pi m\}}\int\limits_{\{|v|\leq\pi m\}} \mathcal{F}f(-u)\mathcal{F}f(v)$$
$$\frac{1}{\Delta^2}\,\mathbb{E}\left[\left(\frac{\widehat{\varphi}_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\varphi'_{\Delta}(u)}{\varphi_{\Delta}(u)}\right)\left(\frac{\widehat{\varphi}_{\Delta,n}(-v)}{\widetilde{\varphi}_{\Delta,n}(-v)}-\frac{\varphi'_{\Delta}(-v)}{\varphi_{\Delta}(-v)}\right)\right]\mathrm{d}u\,\mathrm{d}v$$
$$\leq \frac{CT^{-1}}{(2\pi)^2}\Bigg\{\left(\int|\Psi''(x)|\,\mathrm{d}x+2\int|\Psi'(x)|^2\,\mathrm{d}x\right)\int\limits_{\{|u|\leq\pi m\}}\left|\frac{\mathcal{F}f(-u)}{\varphi_{\Delta}(u)}\right|^2\mathrm{d}u$$
$$\left(\sup_{x\in\mathbb{R}}|\Psi''(x)|+2\sup_{x\in\mathbb{R}}|\Psi'(x)|^2\right)\left(\int\limits_{\{|u|\leq\pi m\}}\left|\frac{\mathcal{F}f(-u)}{\varphi_{\Delta}(u)}\right|\mathrm{d}u\right)^2\Bigg\}. \quad (2.5.30)$$

Using the fact that we have now, by assumption, $\mathcal{F}f\in\mathrm{L}^1(\mathbb{R})$, we can apply Fubini's theorem to find that

$$\mathbb{E}\left[\left|\frac{1}{2\pi}\int\limits_{\{|u|>\pi m\}} \mathcal{F}f(-u)\frac{1}{\Delta}\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\varphi'_{\Delta}(u)}{\varphi_{\Delta}(u)}\right)\mathrm{d}u\right|^2\right]$$
$$= \frac{1}{(2\pi)^2}\int\limits_{\{|u|>\pi m\}}\int\limits_{\{|v|>\pi m\}} \Bigg(\mathcal{F}f(-u)\mathcal{F}f(v)$$
$$\frac{1}{\Delta^2}\,\mathbb{E}\left[\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}-\frac{\varphi'_{\Delta}(u)}{\varphi_{\Delta}(u)}\right)\left(\frac{\widehat{\varphi}'_{\Delta,n}(-v)}{\widetilde{\varphi}_{\Delta,n}(-v)}-\frac{\varphi'_{\Delta}(-v)}{\varphi_{\Delta}(-v)}\right)\right]\Bigg)\mathrm{d}u\,\mathrm{d}v. \quad (2.5.31)$$

An application of Lemma 2.5.1 gives

$$\frac{1}{(2\pi)^2}\int\limits_{\{|u|>\pi m\}}\int\limits_{\{|v|>\pi m\}} \Bigg(\mathcal{F}f(-u)\mathcal{F}f(v)$$

$$\frac{1}{\Delta^2}\,\mathbb{E}\left[\left(\frac{\widehat{\varphi}'_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)} - \frac{\varphi'_{\Delta}(u)}{\varphi_{\Delta}(u)}\right)\left(\frac{\widehat{\varphi}'_{\Delta,n}(-v)}{\widetilde{\varphi}_{\Delta,n}(-v)} - \frac{\varphi'_{\Delta}(-v)}{\varphi_{\Delta}(-v)}\right)\right]\Bigg)\,\mathrm{d}u\,\mathrm{d}v$$

$$\leq \frac{C}{(2\pi)^2}\int\limits_{\{|u|>\pi m\}}\int\limits_{\{|v|>\pi m\}}\Bigg(|\mathcal{F}f(-u)||\mathcal{F}f(v)|$$

$$\Big(\big|\Psi''(u-v)\big| + |\Psi'(u-v)|^2 + \big|\Psi'(u)\big|\,\big|\Psi'(-v)\big|\Big)\Bigg)\,\mathrm{d}u\,\mathrm{d}v. \tag{2.5.32}$$

Finally, arguing again along the same lines as in formulae (2.5.20)-(2.5.24) we find that

$$\frac{C}{(2\pi)^2}\int\limits_{\{|u|>\pi m\}}\int\limits_{\{|v|>\pi m\}}\Bigg(|\mathcal{F}f(-u)||\mathcal{F}f(v)|$$

$$\Big(\big|\Psi''(u-v)\big| + |\Psi'(u-v)|^2 + \big|\Psi'(u)\big|\,\big|\Psi'(-v)\big|\Big)\Bigg)\,\mathrm{d}u\,\mathrm{d}v$$

$$\leq \frac{C}{(2\pi)^2}\Bigg\{\left(\int|\Psi''(x)|\,\mathrm{d}x + 2\int|\Psi'(x)|^2\,\mathrm{d}x\right)\int\limits_{\{|u|>\pi m\}}|\mathcal{F}f(-u)|^2\,\mathrm{d}u$$

$$\wedge\left(\sup_{x\in\mathbb{R}}|\Psi''(x)| + 2\sup_{x\in\mathbb{R}}|\Psi'(x)|^2\right)\Bigg(\int\limits_{\{|u|>\pi m\}}|\mathcal{F}f(-u)|\,\mathrm{d}u\Bigg)^2\Bigg\}. \tag{2.5.33}$$

We have shown that for arbitrary $m \geq 0$,

$$\mathbb{E}\left[\left|\theta - \widehat{\theta}_n\right|^2\right]$$

$$\leq \frac{1}{2\pi^2}\Bigg\{C_1\int\limits_{\{|u|>\pi m\}}|\mathcal{F}f(-u)|^2\,\mathrm{d}u \wedge C_2\left(\int|\mathcal{F}f(-u)|\,\mathrm{d}u\right)^2\Bigg\}$$

$$+\frac{T^{-1}}{2\pi^2}\Bigg\{C_1\int\limits_{\{|u|\leq\pi m\}}\frac{|\mathcal{F}f(-u)|^2}{|\varphi_{\Delta}(u)|^2}\,\mathrm{d}u \wedge C_2\Bigg(\int\limits_{\{|u|\leq\pi m\}}\frac{|\mathcal{F}f(-u)|}{|\varphi_{\Delta}(u)|}\,\mathrm{d}u\Bigg)^2\Bigg\}, \tag{2.5.34}$$

which is the statement of the theorem. □

Again, we can argue exactly along the same lines to see Theorem 2.2.5 and Theorem 2.2.6, using the b) part of Lemma 2.5.1. In the numerator, $\widehat{\varphi}'_{\Delta,n}$ is replaced by $\widehat{\varphi}_{Z_n}$ and one considers $\mathbb{P}^Y$ instead of μ. Moreover $\mathcal{F}\,\mathbb{P}^Y = \varphi_Y$ plays the role of Ψ'' and $|\varphi_Y|^2$ plays the role of $|\Psi'|^2$.

We have to pay attention to the fact that φ_Y is the characteristic function of a probability measure, so the absolute value is bounded by 1. This explains the occurrence of the number 3 instead of $\|\Psi''\|_\infty + 2\|\Psi'\|_\infty^2$.

2.5.2 Proofs of the rate results

Proof of the global rate results

We start by proving the upper bound on the approximation error.

Proof of Lemma 2.3.3. By assumption, f is Sobolev-regular with index s and μ is Sobolev-regular with index $a > -s$. This implies immediately, by duality of Sobolev spaces, that we can pass to the Fourier domain and write

$$\begin{aligned}
&\left| \int f(x)\mu(\,\mathrm{d}x) - \int f(x)(\mathrm{K}_h * \mu)(x)\,\mathrm{d}x \right|^2 \\
= &\left| \frac{1}{2\pi} \int \mathcal{F}f(-u)\mathcal{F}\mu(u)\,\mathrm{d}u - \frac{1}{2\pi} \int \mathcal{F}f(-u)\mathcal{F}\left(\mathrm{K}_h * \mu\right)(u)\,\mathrm{d}u \right|^2 \\
= &\left| \frac{1}{2\pi} \int \mathcal{F}f(-u)(1 - \mathcal{F}\mathrm{K}_h(u))\mathcal{F}\mu(u)\,\mathrm{d}u \right|^2 . \qquad (2.5.35)
\end{aligned}$$

Applying the Cauchy Schwarz inequality and then the regularity assumptions on f and on μ, we find that

$$\begin{aligned}
&\left| \frac{1}{2\pi} \int \mathcal{F}f(-u)(1 - \mathcal{F}\mathrm{K}_h(u))\mathcal{F}\mu(u)\,\mathrm{d}u \right|^2 \\
= &\Bigg| \frac{1}{2\pi} \int \Bigg(\mathcal{F}f(-u)(1 + |u|^2)^{\frac{s}{2}} \\
&\qquad (1 - \mathcal{F}\mathrm{K}_h(u))\mathcal{F}\mu(u)(1 + |u|^2)^{\frac{a}{2}}(1 + |u|^2)^{-\frac{a+s}{2}} \Bigg)\,\mathrm{d}u \Bigg|^2 \\
\leq &\frac{1}{(2\pi)^2} \int |\mathcal{F}f(-u)|^2(1 + |u|^2)^s\,\mathrm{d}u \\
&\qquad \int |1 - \mathcal{F}\mathrm{K}_h(u)|^2(1 + |u|^2)^{-a-s}|\mathcal{F}\mu(u)|^2(1 + |u|^2)^a\,\mathrm{d}u \\
\leq &\frac{1}{(2\pi)^2} \int |\mathcal{F}f(-u)|^2(1 + |u|^2)^s\,\mathrm{d}u \int |\mathcal{F}\mu(u)|^2(1 + |u|^2)^a\,\mathrm{d}u \\
&\qquad \times \sup_{|u|\in\mathbb{R}} |1 - \mathcal{F}\mathrm{K}_h(u)|^2(1 + |u|^2)^{-a-s} \\
\leq &\frac{M_f M_\mu}{(2\pi)^2} \sup_{u\in\mathbb{R}} |1 - \mathcal{F}\mathrm{K}_h(u)|^2(1 + |u|^2)^{-a-s}. \qquad (2.5.36)
\end{aligned}$$

If K is the sinc kernel, we can immediately estimate

$$\sup_{u\in\mathbb{R}} |1 - \mathcal{F}\mathrm{K}(hu)|^2(1 + |u|^2)^{-a-s} = \sup_{|u|\geq\frac{\pi}{h}} (1 + |u|^2)^{-a-s} \leq \pi^{-2a-2s}h^{2a+2s}, \qquad (2.5.37)$$

which gives the desired result.

If $\mathcal{F}\mathrm{K}$ is $\langle a+s\rangle$-times continuously differentiable and the derivatives up to order $\langle a+s\rangle$ at zero vanish, a Taylor series expansion gives for some $\tau \in [-hu, hu]$:

$$\begin{aligned} 1 - \mathcal{F}\mathrm{K}(hu) &= \sum_{k<\langle a+s\rangle} \frac{\mathcal{F}\mathrm{K}^{(k)}(0)}{k!}(hu)^k + \frac{\mathcal{F}\mathrm{K}^{(\langle a+s\rangle)}(\tau)}{\langle a+s\rangle!}(hu)^{\langle a+s\rangle} \\ &= \frac{\mathcal{F}\mathrm{K}^{(\langle a+s\rangle)}(\tau)}{\langle a+s\rangle!}(hu)^{\langle a+s\rangle}. \end{aligned} \tag{2.5.38}$$

Next, Hölder continuity of $\mathcal{F}\mathrm{K}^{\langle a+s\rangle}$ implies that

$$\left|\mathcal{F}\mathrm{K}^{(\langle a+s\rangle)}(\tau)\right| = \left|\mathcal{F}\mathrm{K}^{(\langle a+s\rangle)}(\tau) - \mathcal{F}\mathrm{K}^{(\langle a+s\rangle)}(0)\right| \leq L_{\mathrm{K}}|hu|^{a+s-\langle a+s\rangle}. \tag{2.5.39}$$

Putting (2.5.38) and (2.5.39) together, we have shown that

$$\forall u \in \mathbb{R}: \; |1 - \mathcal{F}\mathrm{K}(hu)| \leq \frac{L_{\mathrm{K}}}{\langle a+s\rangle!}|hu|^{a+s} \wedge (1 + \|\mathcal{F}\mathrm{K}\|_\infty). \tag{2.5.40}$$

From this we conclude that

$$\begin{aligned} &\sup_{u\in\mathbb{R}} |1 - \mathcal{F}\mathrm{K}(hu)|^2 (1+|u|^2)^{-a-s} \\ &\leq \sup_{|u|\leq\frac{\pi}{h}} \frac{L_K^2}{(\langle a+s\rangle!)^2}|hu|^{2a+2s}(1+|u|^2)^{-a-s} \wedge 4 \sup_{|u|>\frac{\pi}{h}} (1+|u|^2)^{-a-s} \\ &\leq \left(\frac{L_K}{\langle a+s\rangle!} + 2/\pi^{a+s}\right)^2 h^{2(a+s)} \end{aligned} \tag{2.5.41}$$

holds true. This gives the desired result thanks to (2.5.35) and (2.5.36). □

The statement of Lemma 2.3.4 is an immediate consequence of the assumptions on the regularity of f and on the decay of φ:

Proof of Lemma 2.3.4. Since $\mathcal{F}\mathrm{K}_h$ is, by assumption, supported on $[-\frac{\pi}{h}, \frac{\pi}{h}]$ and moreover, we trivially have $\|\mathcal{F}\mathrm{K}\|_\infty \leq 1$, we can estimate

$$\begin{aligned} &\int |\mathcal{F}\mathrm{K}(hu)|^2 \left|\frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)}\right|^2 \mathrm{d}u \\ \leq \;& \sup_{|u|\leq\frac{\pi}{h}} \frac{(1+|u|^2)^{-s}}{|\varphi_\Delta(u)|^2} \int |\mathcal{F}f(-u)|^2(1+|u|^2)^s \,\mathrm{d}u. \end{aligned} \tag{2.5.42}$$

The assumptions on the smoothness of f and on the decay of $|\varphi_\Delta|$ give

$$\begin{aligned} &\sup_{|u|\leq\frac{\pi}{h}} \frac{(1+|u|^2)^{-s}}{|\varphi_\Delta(u)|^2} \int |\mathcal{F}f(-u)|^2(1+|u|^2)^s \,\mathrm{d}u \\ \leq \;& M_f C_\varphi \sup_{|u|\leq\frac{\pi}{h}} (1+|u|)^{2\Delta\beta-2s} \exp(2\Delta c_\varphi |u|^\rho). \end{aligned} \tag{2.5.43}$$

On the other hand, we can estimate, using the Cauchy-Schwarz inequality and again the fact that $\mathcal{F}\mathrm{K}_h$ is compactly supported and then the regularity assumption on φ_Δ and f:

$$\begin{aligned}
&\left(\int |\mathcal{F}\,\mathrm{K}(hu)| \left|\frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)}\right| \mathrm{d}u\right)^2 \\
\leq\ & \int\limits_{\{|u|\leq \frac{\pi}{h}\}} \frac{(1+|u|)^{-2s}}{|\varphi_\Delta(u)|^2}\,\mathrm{d}u \int |\mathcal{F}f(-u)|^2(1+|u|^2)^s\,\mathrm{d}u \\
\leq\ & M_f C_\varphi \int\limits_{\{|u|\leq \frac{\pi}{h}\}} (1+|u|)^{2\Delta\beta-2s}\exp\left(2\Delta c_\varphi |u|^\rho\right)\,\mathrm{d}u. \qquad (2.5.44)
\end{aligned}$$

We have thus shown that

$$\begin{aligned}
\sigma_h^2 \ \leq\ & \frac{C_\varphi M_f}{2\pi^2}\Bigg\{C_1 \sup_{|u|\leq\frac{\pi}{h}} (1+|u|)^{2\beta-2s}\exp\left(2\Delta c_\varphi|u|^\rho\right) \\
& \wedge C_2 \int\limits_{\{|u|\leq \frac{\pi}{h}\}} (1+|u|)^{2\Delta\beta-2s}\exp\left(2\Delta c_\varphi |u|^\rho\right)\,\mathrm{d}u\Bigg\}, \qquad (2.5.45)
\end{aligned}$$

which is the statement of Lemma 2.3.4. □

We can now prove the rate results under global regularity assumptions:

Proof of Theorem 2.3.6. Lemma 2.3.4 implies that for $\rho > 0$,

$$v_h \asymp C_1 h^{-2\Delta\beta+2s}\exp\left(\Delta c_\varphi h^{-\rho}\right) \wedge C_2 \int\limits_{\{|u|\leq \frac{\pi}{h}\}} (1+|u|)^{2\Delta\beta-2s}\exp\left(2c_\varphi|u|^\rho\right)\,\mathrm{d}u \qquad (2.5.46)$$

and for $\rho = 0$:

$$v_h \asymp C_1\left(h^{-2\Delta\beta+2s}\vee 1\right) \wedge C_2\left(h^{-2\Delta\beta+2s-1}\vee 1(s=\Delta\beta+1/2)\log(1/h)\vee 1\right). \qquad (2.5.47)$$

Selecting $h^*_{\Delta,n}$ as the minimiser of $b_h + T^{-1}v_h$ gives for $\rho > 0$:

$$h^*_{\Delta,n} \asymp \left(\frac{\log T}{\Delta}\right)^{-\frac{1}{\rho}}. \qquad (2.5.48)$$

For $\rho = 0$ and $C_1 < \infty$, we find that

$$h^*_{\Delta,n} \asymp \begin{cases} 0, & \text{if } s \geq \Delta\beta \\ T^{-\frac{1}{2a+2\Delta\beta}}, & \text{else.} \end{cases} \qquad (2.5.49)$$

For $\rho = 0$ and $C_1 = \infty$, we derive that

$$h^*_{\Delta,n} \asymp \begin{cases} 0, & \text{if } s > \Delta\beta + 1/2 \\ T^{-\frac{1}{2a+2\Delta\beta+1}}, & \text{else} \end{cases} \qquad (2.5.50)$$

We plug $h^*_{\Delta,n}$ in to recover the rates of convergence which are summarised in

Theorem 2.3.6. □

Next, we prove Theorem 2.3.7.

Proof of Theorem 2.3.7. Using again the Sobolev-regularity of f, and the Cauchy-Schwarz inequality, we can estimate for arbitrary $m \geq 0$:

$$\begin{aligned}
& C_1 \int\limits_{\{|u|\geq \pi m\}} |\mathcal{F}f(-u)|^2 \,\mathrm{d}u \wedge C_2 \Bigg(\int\limits_{\{|u|\geq \pi m\}} |\mathcal{F}f(-u)| \,\mathrm{d}u \Bigg)^2 \\
\leq\ & C_1 (\pi m)^{-2s} \int |\mathcal{F}f(-u)|^2 (1+|u|^2)^s \,\mathrm{d}u \\
& \wedge C_2 \int |\mathcal{F}f(-u)|^2 (1+|u|^2)^s \,\mathrm{d}u \int\limits_{\{|u|\geq \pi m\}} (1+|u|^2)^{-s} \,\mathrm{d}u \\
\leq\ & M_f \left(C_1 (\pi m)^{-2s} \wedge C_2 (\pi m)^{-2s+1} \right) =: b_m. \qquad (2.5.51)
\end{aligned}$$

On the other hand, we can argue along the same lines as in the proof of Lemma 2.3.4 to find that for any $m \geq 0$ and for $\rho > 0$,

$$\begin{aligned}
& C_1 \int\limits_{\{|u|\leq \pi m\}} \left| \frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)} \right|^2 \mathrm{d}u \wedge C_2 \Bigg(\int\limits_{\{|u|\leq \pi m\}} \left| \frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)} \right| \mathrm{d}u \Bigg)^2 \\
\lesssim\ & C_1 m^{2\Delta\beta - 2s} \exp\left(2\Delta c_\varphi m^\rho\right) \wedge C_2 \int\limits_{\{|u|\leq \pi m\}} (1+|u|)^{2\Delta\beta-2s} \exp\left(2\Delta c_\varphi |u|^\rho\right) \mathrm{d}u \\
=:\ & v_m. \qquad (2.5.52)
\end{aligned}$$

and for $\rho = 0$,

$$\begin{aligned}
& C_1 \int\limits_{\{|u|\leq \pi m\}} \left| \frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)} \right|^2 \mathrm{d}u \wedge C_2 \Bigg(\int\limits_{\{|u|\leq \pi m\}} \left| \frac{\mathcal{F}f(-u)}{\varphi_\Delta(u)} \right| \mathrm{d}u \Bigg)^2 \\
\lesssim\ & C_1 \left(m^{2\Delta\beta-2s} \vee 1 \right) \wedge C_2 \left(m^{2\Delta\beta-2s+1} \vee 1(s = \Delta\beta + 1/2) \log m \vee 1 \right) \\
=:\ & v_m. \qquad (2.5.53)
\end{aligned}$$

Now, we use the fact that by Theorem 2.2.10, the estimate

$$\begin{aligned}
& \mathbb{E}\left[\left| \theta - \widehat{\theta}_{\Delta,n} \right|^2 \right] \\
\leq\ & \frac{1}{2\pi^2} \Bigg\{ C_1 \int\limits_{\{|u|>\pi m\}} |\mathcal{F}f(-u)|^2 \,\mathrm{d}u \wedge C_2 \left(\int |\mathcal{F}f(-u)| \,\mathrm{d}u \right)^2 \Bigg\} \\
+\ & \frac{T^{-1}}{2\pi^2} \Bigg\{ C_1 \int\limits_{\{|u|\leq \pi m\}} \frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2} \,\mathrm{d}u \wedge C_2 \Bigg(\int\limits_{\{|u|\leq \pi m\}} \frac{|\mathcal{F}f(-u)|}{|\varphi_\Delta(u)|} \,\mathrm{d}u \Bigg)^2 \Bigg\}
\end{aligned}$$

$$\leq \quad b_m + T^{-1} v_m \tag{2.5.54}$$

holds true for *any* $m \geq 0$ to conclude that it holds for the minimiser m^* of $b_m + T^{-1} v_m$. We notice that for $\rho > 0$,

$$m^* \asymp \left(\frac{\log T}{\Delta} \right)^{\frac{1}{\rho}}, \tag{2.5.55}$$

for $\rho = 0$ and $C_1 < \infty$,

$$m^* \asymp \begin{cases} T^{\frac{1}{2\Delta\beta}}, & \text{if } s \leq \Delta\beta \\ \infty, & \text{else} \end{cases} \tag{2.5.56}$$

and for $\rho = 0$ and $C_1 = \infty$,

$$m^* \asymp \begin{cases} T^{\frac{1}{2\Delta\beta}}, & \text{if } s \leq \Delta\beta + \frac{1}{2} \\ \infty, & \text{else.} \end{cases} \tag{2.5.57}$$

From this we readily derive the rate results summarised in Theorem 2.3.7. □

Proof of the local rate result

We start by proving the bound on the approximation error under local regularity assumptions on μ.

Proof of Lemma 2.3.9. We use the trivial observation that the local density g_D can be extended to a compactly supported function $g_1 \in \mathcal{H}_{\mathbb{R}}(a, L', R')$ with constants $L' > L$ and $R' > R$ depending on $d_2 - d_1$, L and R. Let μ_1 be the signed measure with density g_1 and let $\mu_2 := \mu - \mu_1$.

Then the approximation error can be decomposed as follows: Since $\operatorname{supp}(f) = [a, b] \subseteq D$, we have

$$\begin{aligned} & \left| \int f(x)\mu(\,\mathrm{d}x) - \int f(x)(\mathrm{K}_h * \mu)(x)\,\mathrm{d}x \right| \\ = & \left| \int f(x) g_D(x)\,\mathrm{d}x - \int f(x)(\mathrm{K}_h * \mu)(x)\,\mathrm{d}x \right|. \end{aligned} \tag{2.5.58}$$

The decomposition of μ and the triangle inequality imply

$$\begin{aligned} & \left| \int f(x) g_D(x)\,\mathrm{d}x - \int f(x)(\mathrm{K}_h * \mu)(x)\,\mathrm{d}x \right| \\ = & \left| \int f(x)(g_D - \mathrm{K}_h * g_1)(x)\,\mathrm{d}x - \int f(x)(\mathrm{K}_h * \mu_2)(x)\,\mathrm{d}x \right| \\ = & \left| \int f(x)(g_1 - \mathrm{K}_h * g_1)(x)\,\mathrm{d}x - \int f(x)(K_h * \mu_2)(x)\,\mathrm{d}x \right| \\ \leq & \left| \int f(x)(g_1 - \mathrm{K}_h * g_1)(x)\,\mathrm{d}x \right| + \left| \int f(x)(\mathrm{K}_h * \mu_2)(x)\,\mathrm{d}x \right|. \end{aligned} \tag{2.5.59}$$

We start by considering the second summand in the last line of formula (2.5.59).

Recall that f is compactly supported and, by assumption, $|\mathcal{F}f(u)| \leq C_f(1+|u|)^{-s}$. Let $k := -s \vee 0$. Then Theorem B.14 in the appendix tells us that f is a distribution of order k. We can thus estimate

$$\left|\int f(x)\,(\mathrm{K}_h * \mu_2)\,(x)\,\mathrm{d}x\right| \leq \|f\| \sup_{m\leq k}\sup_{x\in[a,b]}\left|(\mathrm{K}_h * \mu_2)^{(m)}(x)\right|. \tag{2.5.60}$$

(For an explanation of what is meant here by $\|f\|$, see Definition B.7.) By assumption, K_h is k-times continuously differentiable, with bounded derivatives. Since, moreover, μ_2 is finite, we can derivate under the integral sign and write

$$\begin{aligned}
&\|f\| \sup_{m\leq k}\sup_{x\in[a,b]}\left|(\mathrm{K}_h * \mu_2)^{(m)}(x)\right| \\
=\;& \|f\| \sup_{m\leq k}\sup_{x\in[a,b]}\left|\left(\mathrm{K}_h^{(m)} * \mu_2\right)(x)\right| \\
=\;& \|f\| \sup_{m\leq k}\sup_{x\in[a,b]}\left|\int h^{-m-1}\,\mathrm{K}^{(m)}\left(\frac{x-y}{h}\right)\mu_2(\,\mathrm{d}y)\right|. \qquad (2.5.61)
\end{aligned}$$

Using the fact that $\operatorname{supp}(f) = [a,b] \subseteq D = (d_1,d_2)$ and that $\mu_2\big|_D \equiv 0$ by construction, we continue from (2.5.61) by estimating

$$\begin{aligned}
&\|f\| \sup_{m\leq k}\sup_{x\in[a,b]}\left|\int h^{-m-1}\,\mathrm{K}^{(m)}\left(\frac{x-y}{h}\right)\mu_2(\,\mathrm{d}y)\right| \\
\leq\;& \|f\| \sup_{m\leq k}\sup_{x\in[a,b]}\left|\int_{\{|x-y|\geq\delta\}} h^{-m-1}\,\mathrm{K}^{(m)}\left(\frac{x-y}{h}\right)\mu_2(\,\mathrm{d}y)\right| \\
\leq\;& \|f\| \sup_{m\leq k}\sup_{z\geq\delta} h^{-m-1}\left|\mathrm{K}^{(m)}\left(\frac{z}{h}\right)\right| |\mu_2|(\mathbb{R}) \qquad (2.5.62)
\end{aligned}$$

with $\delta := (a-d_1)\wedge(d_2-b)$. Finally, the assumptions on the decay of K and its derivatives up to order k give

$$\begin{aligned}
&\|f\| \sup_{m\leq k}\sup_{z\geq\delta} h^{-m-1}\left|\mathrm{K}^{(m)}\left(\frac{z}{h}\right)\right| |\mu_2|(\mathbb{R}) \\
\leq\;& \|f\| C_K |\mu_2|(\mathbb{R})\delta^{-a-(s\vee 0)-1}h^{a+s}. \qquad (2.5.63)
\end{aligned}$$

Next, we have to consider the first summand in the last line of formula (2.5.59).

Consider first the case where f is regular, hence a locally integrable function which has, by assumption, compact support. We use the fact that g_1 is continuous and compactly supported by construction and $\mathrm{K}_h \in \mathrm{L}^1(\mathbb{R})$ by formula (2.3.19) to conclude that

$$\int\int |f(x)||g_1(x-y)\,\mathrm{K}_h(y)|\,\mathrm{d}x\,\mathrm{d}y \leq \|g_1\|_\infty \|\,\mathrm{K}\,\|_{\mathrm{L}^1}\|f\|_{\mathrm{L}^1} < \infty. \tag{2.5.64}$$

We can thus apply Fubini's theorem to find that

$$\int f(x)(g_1 - \mathrm{K}_h * g_1)(x)\,\mathrm{d}x$$

$$= \int f(x) \left(\int g_1(x)\,\mathrm{K}_h(y)\,\mathrm{d}y - \int g_1(x-y)\,\mathrm{K}_h(y)\,\mathrm{d}y \right) \mathrm{d}x$$
$$= \int \mathrm{K}_h(y) \left(\int g_1(x) f(x)\,\mathrm{d}x - \int g_1(x-y) f(x)\,\mathrm{d}x \right) \mathrm{d}y. \quad (2.5.65)$$

If f is non-regular, the identity

$$\int f(x) \int g_1(x-y)\,\mathrm{K}_h(y)\,\mathrm{d}y\,\mathrm{d}x = \int \mathrm{K}_h(y) \int f(x) g_1(x-y)\,\mathrm{d}x\,\mathrm{d}y \quad (2.5.66)$$

remains valid:

In what follows, given a function h, let $\widetilde{h}(x) := h(-x)$. Since we have, by construction, $g_1 \in C^k(\mathbb{R})$ and hence $\mathrm{K}_h * g_1 \in C^k(\mathbb{R})$, $f * \widetilde{g}_1$ as well as $f * (\widetilde{\mathrm{K}}_h * \widetilde{g}_1)$ is well defined as a continuous function.

We can thus write

$$\int f(x) \int g_1(x-y)\,\mathrm{K}_h(y)\,\mathrm{d}y\,\mathrm{d}x = f * (\widetilde{g}_1 * \widetilde{\mathrm{K}}_h)(0). \quad (2.5.67)$$

Now, the convolution operation is associative. This is proved in Theorem 39.3 in [35] for the convolution of distributions with test functions. In Theorem 39.10 the statement is formulated for the convolution of distributions having compact support with infinitely differentiable functions, one of which is compactly supported. The generalization to compactly supported distributions and k-times continuously differentiable functions is straightforward.

We can hence continue from (2.5.67) by writing

$$\int f(x) \int g_1(x-y)\,\mathrm{K}_h(y)\,\mathrm{d}y\,\mathrm{d}x = f * (\widetilde{g}_1 * \widetilde{\mathrm{K}}_h)(0)$$
$$= \widetilde{\mathrm{K}}_h * (f * \widetilde{g}_1)(0) = \int \mathrm{K}_h(y) \int g_1(x-y) f(x)\,\mathrm{d}x\,\mathrm{d}y. \quad (2.5.68)$$

Next, we find that

$$\int \mathrm{K}_h(y) \left(\int g_1(x) f(x)\,\mathrm{d}x - \int g_1(x-y) f(x)\,\mathrm{d}x \right) \mathrm{d}y$$
$$= \int \mathrm{K}_h(y) \left(\int \widetilde{g}_1(0-x) f(x)\,\mathrm{d}x - \int \widetilde{g}_1(y-x) f(x)\,\mathrm{d}x \right) \mathrm{d}y$$
$$= \int \mathrm{K}_h(y) \left(f * \widetilde{g}_1(0) - f * \widetilde{g}_1(y) \right) \mathrm{d}y. \quad (2.5.69)$$

We observe that $f * \widetilde{g}_1$ is $\langle a \rangle + s$-times continuously differentiable and that the derivative of order $\langle a \rangle + s$ is $a - \langle a \rangle$-Hölder continuous.

To see this, we use the fact that $(f * \widetilde{g}_1)^{(\langle a \rangle + s)} = f^{(s)} * \widetilde{g}_1^{\langle a \rangle}$, where $f^{(s)}$ is understood to be a distributional derivative. (See Theorem 41.3 in [35] for explanation.) Now, since $|\mathcal{F} f^{(s)}(-u)| = |u|^s |\mathcal{F} f(-u)| \le C_f$, we can use Theorem B.14 to conclude that $f^{(s)}$ is a compactly supported distribution of order 0. From this, we derive that

$$\left| f^{(s)} * \widetilde{g}_1^{\langle a \rangle}(x) - f^{(s)} * \widetilde{g}_1^{\langle a \rangle}(y) \right|$$

$$
\begin{aligned}
&\leq \|f^{(s)}\| \sup_{\tau\in[a,b]} \left| \widetilde{g}_1^{\langle a\rangle}(x-\tau) - \widetilde{g}_1^{\langle a\rangle}(y-\tau) \right| \\
&\leq \|f^{(s)}\| L' \, |x-y|^{a-\langle a\rangle} . \qquad (2.5.70)
\end{aligned}
$$

The last inequality is valid since $\widetilde{g}_1$ is Hölder-regular by construction.

A Taylor series expansion yields for some $\tau_y \in [0, y]$:

$$
\begin{aligned}
& (f * \widetilde{g}_1)(y) - (f * \widetilde{g}_1)(0) \\
= & \sum_{k<\langle a\rangle+s} \frac{(f*\widetilde{g}_1)^{(k)}(0)}{k!} y^k + \frac{(f*\widetilde{g}_1)^{(\langle a\rangle+s)}(\tau_y)}{(\langle a\rangle+s)!} y^{\langle a\rangle+s} \qquad (2.5.71)
\end{aligned}
$$

Using the assumption on the order of the kernel, we derive from (2.5.71) that

$$
\begin{aligned}
& \left| \int \mathrm{K}_h(y) \left((f*\widetilde{g}_1)(0) - (f*\widetilde{g}_1)(y) \right) \mathrm{d}y \right| \\
= & \left| \sum_{k<\langle a\rangle+s} \frac{(\widetilde{g}_1 * f)^{(k)}(0)}{k!} \int \mathrm{K}_h(y) y^k \, \mathrm{d}y \right. \\
& \left. + \frac{1}{(\langle a\rangle+s)!} \int \mathrm{K}_h(y) \left(g_1 * f\right)^{(\langle a\rangle+s)} (\tau_y) y^{\langle a\rangle+s} \, \mathrm{d}y \right| \\
= & \frac{1}{(\langle a\rangle+s)!} \left| \int \mathrm{K}_h(y) (f*\widetilde{g}_1)^{(\langle a\rangle+s)}(\tau_y) y^{\langle a\rangle+s} \, \mathrm{d}y \right| \\
= & \frac{1}{(\langle a\rangle+s)!} \left| \int \mathrm{K}_h(y) \Big((f*\widetilde{g}_1)^{(\langle a\rangle+s)}(\tau_y) - (f*\widetilde{g}_1)^{(\langle a\rangle+s)}(0) \Big) y^{\langle a\rangle+s} \, \mathrm{d}y \right|. \qquad (2.5.72)
\end{aligned}
$$

Finally, by Hölder-continuity of $(\widetilde{g}_1 * f)^{(\langle a\rangle+s)}$, we can continue from (2.5.72) by estimating

$$
\begin{aligned}
& \frac{1}{(\langle a\rangle+s)!} \left| \int \mathrm{K}_h(y) \Big((f*\widetilde{g}_1)^{(\langle a\rangle+s)}(\tau_y) - (f*\widetilde{g}_1)^{(\langle a\rangle+s)}(0) \Big) y^{\langle a\rangle+s} \, \mathrm{d}y \right| \\
\leq & \frac{1}{(\langle a\rangle+s)!} \|f^{(s)}\| L' \int |\mathrm{K}_h(y)| |y|^{a-\langle a\rangle} |y|^{\langle a\rangle+s} \, \mathrm{d}y \\
= & \frac{1}{(\langle a\rangle+s)!} \|f^{(s)}\| L' \int |\mathrm{K}(z)| |z|^{a+s} \, \mathrm{d}z \, h^{a+s}. \qquad (2.5.73)
\end{aligned}
$$

This completes the proof. □

The proof of Theorem 2.3.12 is now an easy consequence of Lemma 2.3.9.

Proof of Theorem 2.3.12. First, we notice that the upper bounds presented in Lemma 2.3.10 are immediately derived from the assumptions on $\mathcal{F}f$ and on φ_Δ.

Selecting $h^*_{\Delta,n}$ as the minimizer of $b_h + T^{-1} v_h$ gives, for $\rho > 0$:

$$
h^*_{\Delta,n} \asymp \left(\frac{\log T}{\Delta} \right)^{-\frac{1}{\rho}}, \qquad (2.5.74)
$$

for $\rho = 0$ and $C_1 < \infty$:

$$h^*_{\Delta,n} \asymp \begin{cases} T^{-\frac{1}{2\Delta\beta+2a+1}}, & \text{if } s \leq \Delta\beta + 1/2 \\ 0, & \text{else} \end{cases} \tag{2.5.75}$$

and for $\rho = 0$ and $C_1 = \infty$:

$$h^*_{\Delta,n} \asymp \begin{cases} T^{-\frac{1}{2\Delta\beta+2a+2}}, & \text{if } s \leq \Delta\beta + 1/2 \\ 0, & \text{else,} \end{cases} \tag{2.5.76}$$

from which we derive the rate results given in Theorem 2.3.12. □

Chapter 3

Adaptive estimation

In the situation of the preceding chapter, be it in the Lévy model or in the model of density deconvolution, we are given a collection of estimators, indexed by the cutoff parameter m or, more generally, by the bandwidth h and the question which naturally arises in this setting is how to realise the "optimal" choice of the smoothing parameter within the prescribed collection.

We have pointed out that from a point of view which could be called *asymptotic* and *minimax*, an estimator is "good" if, for some given subset $\mathcal{S}$ of the parameter space, the maximal risk over $\mathcal{S}$ is asymptotically close to the minimax risk.

The obvious drawback of this approach is the dependence on the given subset $\mathcal{S}$, say, on the collection of measures which belong (locally) to some prescribed Sobolev or Hölder space.

Ideally, one should aim at constructing estimators which are adaptive minimax, that is, simultaneously asymptotically minimax over many subsets of the parameter space.

The design of minimax adaptive estimators of linear functionals has been widely studied in the literature, starting from the work by Lepski in the early 90s [46, 47]. See Tsybakov [69], Tsybakov and Klemelä [71], Goldenshluger [32], Goldenshluger and Pereverzev [31] and Cai and Low [10, 11].

In the present section, the choice of the smoothing parameter is realised, using model selection techniques, thus following the approach which has been initiated in a series of papers by Massart and Birgé in the late 90s and early 00s, see [6, 53] and [5].

The model selection point of view essentially differs from the methods in the spirit of Lepski in the sense that the problem is considered from a non-asymptotic perspective. Rather than comparing the performance of an estimator with rate optimal estimators over suitable subsets of the parameter space, it is the concern of model selection strategies to compare the risk of an estimator with the optimal risk in some given family of estimators.

Let a collection $(\widehat{\theta}_m)_{m\in\mathcal{M}}$ of estimators be given. Now, the question is: Is it possible to select an "nearly optimal" estimator within this collection, that is, to build an estimator $\widetilde{\theta} = \widehat{\theta}_{\widehat{m}}$ such that for some universal positive constant C, $\widetilde{\theta}$ satisfies

$$\mathbb{E}_\theta\left[\mathrm{d}(\theta,\widetilde{\theta})\right] \leq C \inf_{m\in\mathcal{M}} \mathbb{E}_\theta\left[\mathrm{d}(\theta,\widehat{\theta}_m)\right]\ \forall\theta\in\Theta. \tag{3.0.1}$$

Inequalities of the form (3.0.1) are called *oracle inequalities*.

Deconvolution via model selection has been studied, for example, by Comte, Rozenholc and and Taupin [21]. The estimation of linear functionals in the deconvolution model has been investigated by Butucea and Comte, see [9]. In a white noise framework, the estimation of linear functionals via model selection has been investigated by Laurent, Ludeña and Prieur [44].

However, it is not enough, in the present setting, to merely generalise some of the classical and well known results to the Lévy model.

The particularly interesting problem which arises in the present context and which is not quite standard in the literature is the fact that we have to deal with the problem of adaptation to the unknown characteristic function appearing in the denominator. That is, we face here a non-standard problem of model selection with unknown variance.

For recent research on this topic, we are aware of the work by Comte and Lacour [20], dealing with adaptive estimation with unknown variance for density deconvolution with L^2-loss.

Compared to the approach which has been proposed in that paper and also been sketched in [18], we undergo here a change of perspective. Rather than working, in the first step, with a theoretical deterministic penalty term and then introducing, in the second step, a stochastic penalty term and arguing that both quantities are close to each other, we directly approach the stochastic penalty term and give an argument why it makes immediate sense to work with this quantity.

The main technical step to make the model selection procedure work will be an extension of the classical result by Neumann (see Lemma 2.2.2), making the pointwise control on the characteristic function in the denominator uniform on the real line.

The present chapter is organised as follows: In Section 3.1, we give a brief overview of the principles of model selection via penalization. In Section 3.2, we discuss the application to the particular problem of density deconvolution with unknown distribution of the errors and to nonparametric estimation for Lévy processes. We start, in Subsection 3.2.1 by developing some of our ideas for the structurally simpler case of density deconvolution with L^2-loss. In Subsection 3.2.2 we present a more general approach which covers the problem of adaptive estimation of linear functionals in the Lévy model as well as in the deconvolution model.

To keep the discussion intuitive and free from technicalities, the proofs are postponed to Section 3.3.

3.1 The principles of model selection via penalization

We give here a very short overview of the main ideas behind model selection via penalization, following Massart [52] and Birgé [5].

Consider the following framework: One observes a random object X, for example, a random vector or a stochastic process and wishes to infer some quantity $s \in \mathcal{S}$ related to the unknown distribution of X.

In a Gaussian white noise model, one might think about X as a stochastic process, described by the stochastic differential equation

$$\mathrm{d}X(t) = s(t)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\,\mathrm{d}W(t) \tag{3.1.1}$$

and of s as the underlying regression function, $s \in \mathcal{S} = \mathrm{L}^2([0,1])$.

In a density estimation problem, the random quantity X is typically an *i.i.d.* sample $X = X_1, \cdots, X_n \overset{i.i.d.}{\sim} s$ with density $s \in \mathcal{S} = \mathrm{L}^2(\mathbb{R})$.

Let a family of subspaces $(S_m)_{m\in\mathcal{M}}$ of the original parameter space, called *models* be given. Each S_m reflects some idea about the true nature of the underlying s. To each S_m corresponds some estimator $\widehat{s}_m$ of s. Now, it is the aim of model selection procedures to select the "best" model within the given list.

This is best understood by considering minimum contrast estimators. If γ_n is some empirical criterion based on the observation X, and

$$s = \operatorname*{argmin}_{t\in\mathcal{S}} \mathbb{E}\left[\gamma_n(t)\right], \tag{3.1.2}$$

the *minimum contrast estimator* related to S_m is

$$\widehat{s}_m = \operatorname*{argmin}_{t\in S_m} \gamma_n(t). \tag{3.1.3}$$

The associated natural loss function is $\mathrm{d}(s,t) = \mathbb{E}[\gamma_n(t)] - \mathbb{E}[\gamma_n(s)]$.

Now, here arises the classical trade-off between the approximation error and the error within the model. Choosing a "large model" will allow a good approximation of the underlying function s, thus reducing the bias term, but leads to a large variance. Contrarily, a relatively small model reduces the variance of the estimator, but may be far from the true s.

This point can be illustrated by looking at the white noise model (3.1.1). Let finite dimensional subspaces $(S_m)_{m\in\mathcal{M}}$ of the $\mathrm{L}^2([0,1])$ be given and consider, for each $S_m = \operatorname{span}\{\psi_{m,1}, \cdots, \psi_{m,k}\}$ the projection estimator

$$\widehat{s}_m := \sum_{j=1}^{k} \widehat{c}_{m,j}\psi_{m,j} \tag{3.1.4}$$

with empirical coefficients

$$\widehat{c}_{m,j} = \int \psi_{m,j}(t)\,\mathrm{d}X(t). \tag{3.1.5}$$

The choice of a high dimensional subspace S_m allows a better approximation of s and thus reduces the bias, but leads to a large variance. Choosing S_m small reduces the variance, but the true s may be far from S_m. The oracle choice $m^* \in \mathcal{M}$, which best balances the bias and the variance term depends on the unknown object s and is hence unknown.

In order to construct a parameter $\widehat{m}$ for which the corresponding risk is close to the oracle risk, one considers penalized criteria

$$\gamma_n(\widehat{s}_m) + \mathrm{pen}(m) \tag{3.1.6}$$

and chooses

$$\widehat{m} = \operatorname*{argmin}_{m \in \mathcal{M}} \gamma_n(\widehat{s}_m) + \mathrm{pen}(m). \tag{3.1.7}$$

Then $\widehat{s}_{\widehat{m}}$ satisfies for each $m \in \mathcal{M}$ and any $s_m \in S_m$:

$$\begin{aligned} \mathrm{d}\left(s, \widehat{s}_{\widehat{m}}\right) \leq\ & \mathrm{d}(s, s_m) + \mathrm{pen}(m) \\ +\ & \bar{\gamma}_n(s_m) - \bar{\gamma}_n(\widehat{s}_{\widehat{m}}) - \mathrm{pen}(\widehat{m}), \end{aligned} \tag{3.1.8}$$

with

$$\bar{\gamma}_n(t) := \gamma_n(t) - \mathbb{E}[\gamma_n(t)]. \tag{3.1.9}$$

The penalty should be chosen small enough to guarantee that, for any $m \in \mathcal{M}$, $\mathrm{d}(s, s_m) + \mathrm{pen}(m)$ is not much larger than $\mathbb{E}[\,\mathrm{d}(s, \widehat{s}_m)]$. On the other hand, $\mathrm{pen}(m)$ should be large enough to bound, with high probability, the fluctuation of $\bar{\gamma}_n(s_m) - \bar{\gamma}_n(\widehat{s}_{\widehat{m}})$, leading to an oracle inequality

$$\mathbb{E}[\,\mathrm{d}(s, \widehat{s}_{\widehat{m}})] \leq C \inf_{m \in \mathcal{M}} \mathbb{E}[\,\mathrm{d}(s, \widehat{s}_m)] \tag{3.1.10}$$

with some universal constant C.

The idea to select an appropriate model by looking at penalized criteria has already been considered in the early 70s by Akaike [1] or Mallows [50].

However, the precise control on the empirical process $\bar{\gamma}_n(s) - \bar{\gamma}_n(t)$ is, in many cases, a highly non-trivial issue. The main technical tools involved in such arguments are non-asymptotic concentration inequalities for empirical processes, which have been initiated about 15 years ago by Talagrand [66, 67] and further developed, for example, by Ledoux [45], Massart [51], Rio [60, 61], Klein [41], Bousquet [7] and Klein and Rio [42].

For a detailed discussion on the role of concentration inequalities in model selection, we refer to Massart [52].

3.2 Adaptive estimation for Lévy processes and for density deconvolution with unknown distribution of the errors

We consider model selection techniques in the model of nonparametric deconvolution with unknown distribution of the noise as well as in the Lévy model.

Although the focus of this thesis clearly lies on the Lévy model, we believe that the results obtained for density deconvolution are interesting in their own right so it makes sense to discuss both concepts in parallel.

In fact, the ideas developed here are fairly general and it does not make much of a difference weather we talk about Lévy processes or density deconvolution, about estimating linear functionals or about L^2-loss.

Let us briefly recall the setting: In the deconvolution model, we observe

$$Z_j = Y_j + \varepsilon_j,\ j = 1, \cdots, n, \tag{3.2.1}$$

where (Y_j) and (ε_j) are independent sequences of i.i.d. random variables. The Y_j are distributed according to $\mathbb{P}^Y$ and the ε_j are distributed according to f_ε, which is unknown. We assume that n *i.i.d.* copies $\varepsilon_{-n}, \cdots, \varepsilon_{-1} \sim f_\varepsilon$ of the pure noise are available.

In the Lévy model, we observe

$$X_\Delta, \cdots, X_{2n\Delta}, \tag{3.2.2}$$

where X is a Lévy process having no drift and no Gaussian component, finite second moments and finite variation on compact sets.

The goal is to estimate some linear functional $\theta = \langle f, \mathbb{P}^Y \rangle$ of the unknown distribution of the Y_j or some functional $\theta = \langle f, \mu \rangle$ of the finite signed measure $\mu = x\nu(\,\mathrm{d}x)$.

For notational convenience, we suppress in this chapter, the dependence on the observation distance Δ and assume, without loss of generality, that $\Delta = 1$. So it is understood in the sequel, that $\varphi = \varphi_1$ and $\varphi' = \varphi_1'$ and the same notation is used when talking about the empirical versions.

Passing to the Fourier domain and replacing the unknown characteristic functions by their (truncated) empirical versions an then applying some kernel smoothing procedure leads to a collection of estimators, indexed by the bandwidth. We apply model selection techniques to realise the "optimal" data driven choice of the bandwidth, thus adapting automatically to the unknown smoothness of $\mathbb{P}^Y$ or μ and to the unknown decay of the characteristic function φ.

Adaptive estimation of linear functionals via model selection has been considered, for example, in [44] and [9]. The crucial difference is that, in the present situation, one has to deal with the additional complication that the variance term in unknown.

In the setting of adaptive estimation for Lévy processes, Comte and Genon-Catalot [18] have dealt with this problem by proposing an additional a priori assumption on the size of the model. However, this assumption is critical since it depends itself on the unknown smoothness parameters.

Recently, Comte and Lacour [20] have proposed a strategy to dealing with the unknown variance term which does not rely on any prior knowledge of the smoothness parameter. However, this strategy is designed for L^2-loss and spectral cutoff estimators and the generalisation to the estimation of linear functionals with general kernels is not straightforward. Moreover, it is assumed in that paper that the feasible number M of observations of the pure noise is, by a polynomial factor, of larger order than the number n of observations of the noisy data and would lead to a loss of polynomial order when assuming that $M = n$. Since we aim at generalising our approach to the Lévy setting where one has naturally to assume that $M = n$, we propose a different strategy and can avoid the loss of a polynomial factor.

3.2.1 A first approach to model selection with unknown variance: Density deconvolution with L^2-loss

For the moment, we leave Lévy processes aside and place ourselves in the model of density deconvolution.

We start by treating the structurally simpler case were $\mathbb{P}^Y$ has a square integrable Lebesgue density g and one is interested in L^2-loss rather than in estimating linear functionals. Moreover, we limit our considerations to smoothing with the sinc-kernel.

Let us first have a look at the setting where the distribution of the errors is known, and which is more standard in the literature on model selection, see for example Comte et al. [21].

For $m \in \mathbb{N}$, the corresponding spectral cutoff estimator of g is given by

$$\widehat{g}_{m,n} := \mathcal{F}^{-1}\left(\frac{\widehat{\varphi}_{Z_n}(u)}{\varphi_\varepsilon(u)} 1\left([-\pi m, \pi m]\right)(u)\right), \tag{3.2.3}$$

with $\widehat{\varphi}_{Z_n}$ defined as in formula (2.2.4). Moreover,

$$\begin{aligned} g_m &:= \mathcal{F}^{-1}\Big((\mathcal{F}g)(u) 1\left([-\pi m, \pi m]\right)(u)\Big) \\ &= \mathcal{F}^{-1}\left(\left(\frac{\varphi_Z(u)}{\varphi_\varepsilon(u)}\right) 1\left([-\pi m, \pi m]\right)(u)\right) \end{aligned} \tag{3.2.4}$$

denotes the projection of g on the space of square integrable functions with Fourier transform supported in $[-\pi m, \pi m]$.

The L^2-risk of $\widehat{g}_{m,n}$ can be estimated as follows:

$$\begin{aligned} \mathbb{E}\left[\|g - \widehat{g}_{m,n}\|_{\mathrm{L}^2}^2\right] &= \|g - g_m\|_{\mathrm{L}^2}^2 + \mathbb{E}\left[\|g_m - \widehat{g}_m\|_{\mathrm{L}^2}^2\right] \\ &= \|g - g_m\|_{\mathrm{L}^2}^2 + \frac{n^{-1}}{2\pi} \int\limits_{\{|u| \leq \pi m\}} \frac{\mathbb{E}\left[|\varphi_Z(u) - \widehat{\varphi}_{Z_1}(u)|^2\right]}{|\varphi_\varepsilon(u)|^2} \mathrm{d}u \\ &\leq \|g - g_m\|_{\mathrm{L}^2}^2 + \frac{n^{-1}}{2\pi} \int\limits_{\{|u| \leq \pi m\}} \left|\frac{1}{\varphi_\varepsilon(u)}\right|^2 \mathrm{d}u \\ &=: \|g - g_m\|_{\mathrm{L}^2}^2 + n^{-1}\sigma_m^2. \end{aligned} \tag{3.2.5}$$

Suppose that we are given a collection $\mathcal{M} := \{1, \cdots, m_n\}$ of cutoff parameters. For L^2-estimation, one will typically assume that $m_n = n$.

Ideally, the cutoff parameter m should be chosen to realise the best trade-off between the bias term and the quantity σ_m^2 which bounds the variance. However, this ideal choice is not feasible and can only serve as a benchmark since it depends on the unknown g.

Pythagoras' theorem yields the following decomposition of the bias term:

$$\|g - g_m\|_{\mathrm{L}^2}^2 = \|g\|_{\mathrm{L}^2}^2 - \|g_m\|_{\mathrm{L}^2}^2. \tag{3.2.6}$$

Consequently, the problem of realising the optimal bias-variance trade-off can be considered independent of the constant summand $\|g\|_{\mathrm{L}^2}^2$. Next, $\|g_m\|_{\mathrm{L}^2}^2$ can be estimated from the data by the natural bias corrected version $\|\widehat{g}_{m,n}\|_{\mathrm{L}^2}^2 - n^{-1}\sigma_m^2$. (Indeed, this is an overcorrection, since σ_m^2 is an upper bound for the variance.)

However, since this will not give control on the fluctuation simultaneously over countably many m, σ_m^2 is replaced by $\lambda_m\sigma_m^2$ with logarithmic weights λ_m chosen such that $\sum\limits_{m\in\mathcal{M}} \sigma_m^2 e^{-\lambda_m} < \infty$.

These considerations lead to choosing the cutoff parameter $\widehat{m}$ as the minimiser of the penalized criterion

$$-\|\widehat{g}_m\|_{\mathrm{L}^2}^2 + \mathrm{pen}(m) \tag{3.2.7}$$

with penalty term $\mathrm{pen}(m) = 2n^{-1}\lambda_m\sigma_m^2$.

So far, the discussion crucially depends on the fact that one can control the fluctuation of certain *stochastic* terms by penalizing with some *deterministic* quantity.

Now the quantity σ_m^2 involved in the definition of the penalty term depends on the characteristic function φ_ε of the noise, which is itself assumed to be unknown in the present setting. This is the point where our analysis starts.

The same subject has been treated by Comte and Lacour, see [20]. However, our approach is slightly different.

In the sequel, we set

$$\widehat{g}_{m,n} := \mathcal{F}^{-1}\left(\frac{\widehat{\varphi}_{Z_n}(u)}{\widetilde{\varphi}_{\varepsilon_n}(u)} 1\left([-\pi m, \pi m]\right)(u)\right), \tag{3.2.8}$$

with $\widehat{\varphi}_{Z_n}$ defined as in formula (2.2.4) and with $\frac{1}{\widetilde{\varphi}_{\varepsilon_n}}$ defined as in formula (2.2.6) and

$$\widetilde{g}_m := \mathcal{F}^{-1}\left(\frac{\varphi_Z}{\widetilde{\varphi}_{\varepsilon_n}} 1\left([-\pi m, \pi m]\right)(u)\right). \tag{3.2.9}$$

The squared risk can be estimated as follows:

$$\begin{aligned}
\mathbb{E}\left[\|g - \widehat{g}_{m,n}\|_{\mathrm{L}^2}^2\right] &\leq \|g - g_m\|_{\mathrm{L}^2}^2 + C\frac{n^{-1}}{2\pi} \int\limits_{\{|u|\leq \pi m\}} \left|\frac{1}{\varphi_\varepsilon(u)}\right|^2 \mathrm{d}u \\
&=: \|g - g_m\|_{\mathrm{L}^2}^2 + n^{-1}\sigma_m^2. \qquad (3.2.10)
\end{aligned}$$

with constant $C := 8C_2 + 2$, where C_2 denotes the constant in Neumann's Lemma. For the proof of this estimate, see [20]. (To avoid confusion about the role of M, let us recall at this point that we treat here, in contrast to the situation in [20], the special case where the number M of observations of the pure noise is equal to the number n of observations of the noisy data.)

At the first glance, it is well intuitive to redefine the penalty term, replacing the unknown $\frac{1}{\varphi_\varepsilon}$ by $\frac{1}{\widetilde{\varphi}_{\varepsilon_n}}$, defined as in formula (2.2.6). This approach leads to considering a stochastic version of the penalty term:

$$\widetilde{\mathrm{pen}}(m) = \gamma n^{-1}\widetilde{\lambda}_m\widetilde{\sigma}_m^2, \tag{3.2.11}$$

with some purely numerical constant γ, with

$$\widetilde{\sigma}_m^2 := C\frac{1}{2\pi}\int\limits_{\{|u|\leq \pi m\}}\left|\frac{1}{\widetilde{\varphi}_\varepsilon(u)}\right|^2\,\mathrm{d}u \tag{3.2.12}$$

and with logarithmic weights $\widetilde{\lambda}_m$ to be appropriately chosen.

One can argue that, conditional on $\widehat{\varphi}_{\varepsilon_n}$, $\widetilde{\mathrm{pen}}(m)$ is appropriate for giving uniform control on the fluctuation of

$$\|\widehat{g}_{m,n}\|_{\mathrm{L}^2}^2 = \frac{1}{2\pi}\int\limits_{\{|u|\leq \pi m\}}\left|\frac{\widehat{\varphi}_{Z_n}(u)}{\widetilde{\varphi}_{\varepsilon_n}(u)}\right|^2\,\mathrm{d}u \tag{3.2.13}$$

round its target.

However, this will a priori only guarantee that the following holds true with high probability:

$$\|\widehat{g}_{m,n}\|_{\mathrm{L}^2}^2 - \widetilde{\mathrm{pen}}(m) \leq \|\widetilde{g}_m\|_{\mathrm{L}^2}^2 \quad \forall m\in\mathcal{M}. \tag{3.2.14}$$

But nothing warrants that with with high probability,

$$\|\widetilde{g}_m\|_{\mathrm{L}^2}^2 \leq \|g_m\|_{\mathrm{L}^2}^2 \quad \forall m\in\mathcal{M}. \tag{3.2.15}$$

The quantity $\widetilde{\sigma}_m^2$ involved in the definition of the stochastic penalty term may be far from the true σ_m^2 and moreover,

$$\|\widetilde{g}_m - g_m\|_{\mathrm{L}^2}^2 = \frac{1}{2\pi}\int\limits_{\{|u|\leq \pi m\}}|\varphi_Z(u)|^2\left|\frac{1}{\widetilde{\varphi}_\varepsilon} - \frac{1}{\varphi_\varepsilon}\right|^2\,\mathrm{d}u \tag{3.2.16}$$

may be large.

Ideally, one should subtract another deterministic correction term in order to bound the fluctuation of $\frac{1}{\widetilde{\varphi}_{\varepsilon_n}}$ around $\frac{1}{\varphi_\varepsilon}$. But again, this would only be possible if $\frac{1}{\varphi_\varepsilon}$ were feasible.

Still, on an intuitive level, one might argue as follows: Close to the origin, the characteristic function φ_ε is relatively large and $\widehat{\varphi}_{\varepsilon_n}$ is reliable as an estimator of φ_ε. So there is some hope that replacing $\frac{1}{\varphi_\varepsilon}$ by $\frac{1}{\widehat{\varphi}_{\varepsilon_n}}$ will not cause a large error there.

Contrarily, once φ_ε is below some critical threshold of order $n^{-1/2}$, the noise is dominant and estimation of this object no longer makes sense. This suggests to terminate the procedure at this stage.

To illustrate this idea in more detail: Assume that for some $U_n\in\mathbb{R}$ and some constant C'_φ, we have

$$\forall u\in[-U_n,U_n]^c: \quad |\varphi_\varepsilon(u)| \leq C'_\varphi n^{-1/2}. \tag{3.2.17}$$

Let indices k and m with $\pi k > \pi m \geq U_n$ be given. Now, when using

$$\int\limits_{\{\pi m \leq |u| \leq \pi k\}} \left| \frac{1}{\widetilde{\varphi}_{\varepsilon_n}(u)} \right|^2 \mathrm{d}u \tag{3.2.18}$$

as an estimator of

$$\int\limits_{\{\pi m \leq |u| \leq \pi k\}} \left| \frac{1}{\varphi_\varepsilon(u)} \right|^2 \mathrm{d}u, \tag{3.2.19}$$

we simply accumulate pure noise in formula (3.2.18) while the quantity in (3.2.19) blows up. Hence, using $\widetilde{\sigma}_k^2 - \widetilde{\sigma}_m^2$ as an estimator of the true $\sigma_k^2 - \sigma_m^2$ makes no sense for indices larger than $\frac{1}{\pi} U_n$.

On the other hand: Does it make sense at all to consider such indices? Clearly, we can estimate for $\pi m, \pi k \geq U_n$:

$$\begin{aligned}
\|g_k - g_m\|_{\mathrm{L}^2}^2 &= \frac{1}{2\pi} \int\limits_{\{\pi m \leq |u| \leq \pi k\}} |\mathcal{F}g(u)|^2 \,\mathrm{d}u \\
&= \frac{1}{2\pi} \int\limits_{\{\pi m \leq |u| \leq \pi k\}} |\varphi_\varepsilon(u)|^2 \left| \frac{\mathcal{F}g(u)}{\varphi_\varepsilon(u)} \right|^2 \mathrm{d}u \\
&\leq \left(C'_\varphi\right)^2 \frac{n^{-1}}{2\pi} \int\limits_{\{\pi m \leq |u| \leq \pi k\}} \left| \frac{1}{\varphi_\varepsilon(u)} \right|^2 \mathrm{d}u \\
&\leq \left(C'_\varphi\right)^2 \left(\sigma_k^2 - \sigma_m^2\right).
\end{aligned} \tag{3.2.20}$$

From this we conclude that, when considering indices larger than $\frac{1}{\pi} U_n$, the gain which results form reducing the approximation error is (up to a constant factor) smaller than the unavoidable loss which is caused by the larger variance term.

These heuristics lead to proposing the following strategy: Estimation of the unknown characteristic function in the denominator is restricted to some compact set where φ_ε is large enough and a reasonable estimator is hence feasible. Outside this set, the estimator is set to zero. Of course, this global thresholding must be formulated in terms of $\widehat{\varphi}_{\varepsilon_n}$, so the compact set is in fact random. To avoid terminating too early or too late, we introduce an extra logarithmic factor in the thresholding scheme, which allows to derive exponential inequalities.

This leads to the following alternative definition of an estimator of $\frac{1}{\varphi} := \frac{1}{\varphi_\varepsilon}$:

Definition 3.2.1. *For some $\delta > 0$, let the weight function w be defined by*

$$w(u) := (\log(e + |u|))^{-\frac{1}{2} - \delta}. \tag{3.2.21}$$

For some positive constants κ to be chosen, set

$$\frac{1}{\check{\varphi}_n(u)} := \frac{1}{\check{\varphi}_n^\kappa(u)} := \frac{\mathbb{1}\left(\left\{\forall t \in [-u, u] : |\widehat{\varphi}_n(t)| \geq \kappa \,(\log n)^{1/2}\, w(t)^{-1} n^{-1/2}\right\}\right)}{\widehat{\varphi}_n(u)}. \tag{3.2.22}$$

The definition of the weight function is originally due to Reiß and Neumann and the results obtained in [56] will play an important role for our arguments.

The above definition is meaningful under the following rather general assumption on the characteristic function:

Assumption 3.2.2. *There is a function h which is non-decreasing on $\mathbb{R}^-$ and non-increasing on $\mathbb{R}^+$ and positive constants C_φ, C'_φ such that*

$$\forall u \in \mathbb{R}: \ C_\varphi h(u) \leq |\varphi(u)| \leq C'_\varphi h(u). \tag{3.2.23}$$

Now, we can and will argue as follows: On the random compact set before the estimator is set to zero, $\frac{1}{\tilde{\varphi}_n}$ is sufficiently close to $\frac{1}{\varphi}$ not only pointwise, but uniformly. This makes the model selection procedure work as if $\frac{1}{\varphi}$ were feasible.

On the other hand, the argument given in (3.2.20) will tell us that we do not run the risk of throwing away indices which would have been needed for the model selection procedure.

We proceed as follows: We redefine $\widehat{g}_{m,n}$ in terms of $\frac{1}{\check{\varphi}_n(u)}$ and set, in the sequel:

$$\widehat{g}_{m,n} := \mathcal{F}^{-1}\left(\frac{\widehat{\varphi}_{Z_n}(u)}{\check{\varphi}_n(u)} 1\left([-\pi m, \pi m]\right)(u)\right). \tag{3.2.24}$$

The stochastic penalty term is defined by

$$\widetilde{\mathrm{pen}}(m) := 200\left(c_1^2 \widetilde{\lambda}_m^2 + \kappa^2 \log n\right) n^{-1} \widetilde{\sigma}_m^2, \tag{3.2.25}$$

where c_1 is the constant appearing in Talagrand's inequality (see Lemma 3.3.3) and

$$\widetilde{\sigma}_m^2 := \frac{1}{2\pi} \int\limits_{\{|u| \leq \pi m\}} \left|\frac{w(u)^{-1}}{\check{\varphi}_n(u)}\right|^2 \mathrm{d}u. \tag{3.2.26}$$

Finally, the weights are defined to be

$$\widetilde{\lambda}_m := \log\left(\widetilde{\sigma}_m^2 m^2\right). \tag{3.2.27}$$

This leads to defining the random cutoff parameter $\widehat{m}$ as follows:

Definition 3.2.3. *Let $\mathcal{M}_n = \{1, \cdots, n\}$ be given. Let $\widehat{g}_{m,n}$ be defined by (3.2.24) and let $\widetilde{\mathrm{pen}}(m)$ be defined as in (3.2.25). Then we set*

$$\begin{aligned}\widehat{m} &:= \operatorname{argmin}\left\{m \in \mathcal{M} : -\|\widehat{g}_m\|_{\mathrm{L}^2}^2 + \widetilde{\mathrm{pen}}(m)\right\} \\ &:= \min\left\{m \in \mathcal{M} : -\|\widehat{g}_m\|_{\mathrm{L}^2}^2 + \widetilde{\mathrm{pen}}(m) \leq -\|\widehat{g}_j\|_{\mathrm{L}^2}^2 + \widetilde{\mathrm{pen}}(j) \ \forall j \in \mathcal{M}\right\}.\end{aligned} \tag{3.2.28}$$

We can now formulate the main result of this subsection:

Theorem 3.2.4. *Let observations according to Model 2 be given. Let the assumptions (D1) and (D3) be satisfied. Assume that $\mathbb{P}^Y$ has a square integrable*

Lebesgue density g. For $m \in \mathbb{N}$, let

$$\sigma_{w,m}^2 := \frac{1}{2\pi} \int\limits_{\{|u| \leq \pi m\}} \left| \frac{w(u)^{-1}}{\varphi_\varepsilon(u)} \right|^2 du \tag{3.2.29}$$

and

$$\lambda_m := \sqrt{2} \left(\log \left(\sigma_{w,m} m \right) \right)^{1/2}. \tag{3.2.30}$$

Moreover, let

$$\mathrm{pen}(m) := 200 \left(c_1^2 \lambda_m^2 + \kappa^2 \log n \right) \sigma_{w,m}^2. \tag{3.2.31}$$

Let $\widehat{g}_{m,n}$ be defined according to formula (3.2.24). Assume that for some $\gamma > 0$, $\kappa \geq 2 \left(\sqrt{4c_1} + \gamma \right)$. Let Assumption 3.2.2 be satisfied. Let a collection of indices $\mathcal{M} := \{m_1, \cdots, m_n\}$ be given. Let $\widehat{m}$ be the random cutoff parameter introduced in Definition 3.2.3. Then we can give the following bound on the corresponding L^2*-risk:*

$$\mathbb{E} \left[\| g - \widehat{g}_{\widehat{m}} \|_{\mathrm{L}^2}^2 \right] \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ \| g - g_m \|_{\mathrm{L}^2}^2 + \mathrm{pen}(m) \right\} + C' n^{-1},$$

with some constant C' depending only on the choice of κ and δ and a constant C^{ad} depending on the choice of κ and δ and on the ratio C_φ'/C_φ.

Some remarks are in order: It is interesting to note that our reasoning has some structural similarities with the approach proposed in [18] as well as with the approach which has been considered in [20].

In [18], Comte and Genon-Catalot work under the a priori assumption that for some nonnegative index β and positive constants C_φ and C_φ', the characteristic function has a polynomial decay behaviour,

$$\forall u \in \mathbb{R}: \; C_\varphi (1 + |u|^2)^{-\frac{\beta}{2}} \leq |\varphi(u)| \leq C_\varphi' (1 + |u|^2)^{-\frac{\beta}{2}}. \tag{3.2.32}$$

The quasi-monotonicity condition which we formulate in Assumption 3.2.2 can be understood as a generalisation of (3.2.32).

Moreover, the a priori assumption that for the size m_n of the collection,

$$\exists \varepsilon \in (0,1), C > 0: \; m_n \leq C n^{(1-\varepsilon)/2\beta} \tag{3.2.33}$$

holds, has been formulated therein.

A similar assumption is *automatically* satisfied in the present setting. This is due to the fact that we automatically terminate at a certain threshold. Our global threshold estimator of the characteristic function leads, if the characteristic function decays polynomially, to considering, with high probability, only indices m for which

$$m \leq C \left(\frac{n}{\log n} \right)^{1/2\beta} \tag{3.2.34}$$

holds true.

The decay behaviour of the characteristic function is no longer assumed to be explicitly known, but comes in implicitly.

In the paper by Comte and Lacour [20], the size of the model is chosen at random and our reasoning can be understood in the same sense. The main difference lies in the fact that we directly approach the characteristic function and introduce an additional logarithmic factor in the thresholding scheme. This is the reason why we can avoid the loss of a polynomial factor.

Contrarily to the assumptions formulated in [20], our reasoning does not rely on any semiparametric assumption on the shape of the characteristic function, such as exponential or polynomial decay behaviour. The only thing which is needed is Assumption 3.2.2, which is fairly general.

It is worth mentioning that we can formulate an analogous result for estimating a Lévy density with L^2-loss.

So far, our reasoning is already fairly general. Surprisingly, we can give some refined argument and treat the above arguments as a special case.

The reason for having discussed the global threshold scheme in detail is the fact that this approach is relevant in applications since one will, anyhow, have to work with a compact approximation of the Fourier transform.

From a theoretical perspective, the reasoning presented in the next section is more satisfactory.

Illustration: Simulation example

The focus of this thesis clearly lies on the theoretical results and not on systematic simulation studies. Still, we want to have a look on some numerical example to see how the estimator performs in applications.

We apply the adaptive estimation procedure to standard normal random variables with $\Gamma(1,2)$ distributed noise, to $\Gamma(3,2)$ random variables with $\Gamma(2,2)$ distribution of the noise, to $\Gamma(5,2)$-random variables with standard normally distributed errors and to $\Gamma(2,2)$ random variables with standard normal distribution of the noise.

The main difficulty which arises in the simulation studies is the fact that the theoretical constants appearing in the definition of the penalty term are in praxis far too large and lead to a bad performance of the estimator.

This problem is dealt with by calibrating the constants in preliminary simulation experiments, which leads to choosing $\kappa = 2$ and working with the effective penalty term

$$\widetilde{\mathrm{pen}}(m) = 4\left(\widetilde{\lambda}_m + (\log n)\right)\widetilde{\sigma}_m^2, \tag{3.2.35}$$

so the constants are in praxis chosen much smaller than the theory would have suggested. Indeed, choosing the constants according to the theoretical results will lead to very undesirable over-penalization effects, especially on small sample sizes. This is not surprising, since the theory is designed to control even highly irregular models, which will hardly ever occur in applications.

Calibrating the effective constants in preliminary simulation experiments is not particular to our reasoning, but quite standard, see, for example, the simulation studies in [19]. Another possibility would be to introduce another sample splitting and use a training set in order to choose the effective constants.

The performance of the adaptive estimator is compared to the "estimator" with oracle choice of the bandwidth. (It would perhaps be more natural to compare the empirical risk of the adaptive estimator directly to the oracle risk. But the oracle risk is not available in a closed form, so it seems reasonable to approximate this quantity by simulation experiments.)

We consider $n = 500, 1000, 10000$ observations and calculate the empirical risk of the adaptive estimator, as well as of the estimator with oracle choice of the bandwidth from 250 independent iterations of the procedure.

The values, multiplied by 100, are collected in the table below.

	$Y_j \sim \mathcal{N}(0,1)$, $\varepsilon_j \sim \Gamma(1,2)$		$Y_j \sim \Gamma(3,2)$, $\varepsilon_j \sim \Gamma(2,2)$	
n	$\widehat{r}_{or}$	$\widehat{r}_{ad}$	$\widehat{r}_{or}$	$\widehat{r}_{ad}$
500	0,26	1,67	1,46	3,13
1000	0,14	0,19	1,07	3,01
10000	0,01	0,13	0,14	0,39
	$Y_j \sim \Gamma(5,2)$, $\varepsilon_j \sim \mathcal{N}(0,1)$		$Y_j \sim \Gamma(2,2)$, $\varepsilon \sim \mathcal{N}(0,1)$	
n	$\widehat{r}_{or}$	$\widehat{r}_{ad}$	$\widehat{r}_{or}$	$\widehat{r}_{ad}$
500	1,55	4,60	10,67	22,45
1000	0,97	4,57	8,82	22,16
10000	0,45	0,48	5,72	9,09

The adaptive procedure performs remarkably well in all cases investigated here, and the most interesting question which arises from the data examples is certainly how the practical choice of the constant can be justified from a theoretical point of view.

3.2.2 A second approach: Estimating linear functionals in the density deconvolution model and in the Lévy model

Let us now turn away from L^2-loss and return to the situation where one is interested in estimating linear functionals. Model selection techniques in this setting have been considered by Laurent, Ludeña and Prieur [44] in a white noise framework, as well as by Butucea and Comte [9] in the deconvolution model with known error distribution.

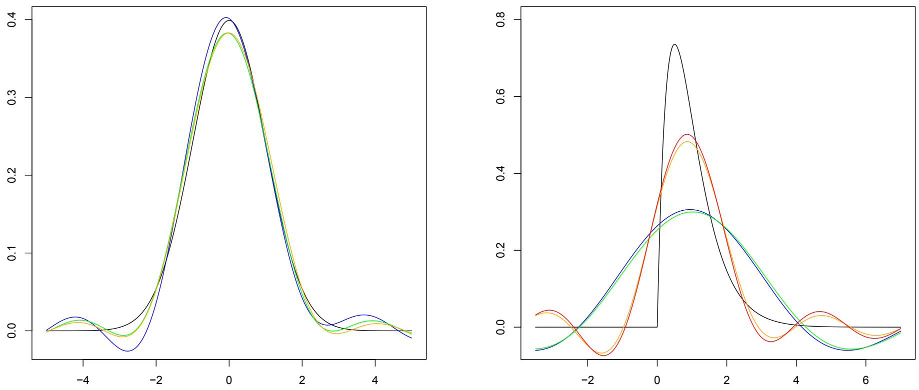

Figure 3.1: The plot on the left shows the true (black) and the estimated probability density, calculated from $n = 500$ (blue), $n = 1000$ (green) and $n = 10000$ (orange) observations for standard normally distributed random variables with $\Gamma(1, 2)$ distribution of the noise. On the right, the performance of the estimator for a relatively non-smooth ($\Gamma(2, 2)$) underlying density (black) with smooth ($\mathcal{N}(0, 1)$) errors is shown. The adaptive estimator is calculated from $n = 500$ (blue), $n = 1000$ (green), $n = 10000$ (orange) and $n = 50000$ (red) observations.

Let a collection

$$\mathcal{M} := \mathcal{M}_n := \{m_1, \cdots, m_n\} \subseteq \mathbb{N} \tag{3.2.36}$$

of indices be given. Let

$$\mathcal{H} := \mathcal{H}_n := \{h_1, \cdots, h_n\} := \left\{\frac{1}{m_1}, \cdots, \frac{1}{m_n}\right\} \tag{3.2.37}$$

be a collection of bandwidths associated with $\mathcal{M}_n$.

For each $m \in \mathcal{M}$, let $\widehat{\theta}_{h_m,n}$ denote the kernel estimator introduced in Definition 2.2.4 or in Definition 2.2.7. Indeed, the discussion for the Lévy model and for density deconvolution is completely analogous.

For notational convenience, we write, in the sequel, $\widehat{\theta}_{m,n}$ instead of $\widehat{\theta}_{h_m,n}$.

In the Lévy model, let for arbitrary $m \in \mathbb{N}$:

$$\begin{aligned} \theta_m &:= \frac{1}{2\pi} \int \mathcal{F}f(-u)\mathcal{F}\mu(u)\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right) \mathrm{d}u \\ &= \int f(x) \left(\mathrm{K}_{h_m} * \mu\right)(x) \, \mathrm{d}x. \end{aligned} \tag{3.2.38}$$

In the convolution model, we write

$$\begin{aligned} \theta_m &:= \frac{1}{2\pi} \int \mathcal{F}f(-u)\mathcal{F}\mathbb{P}^Y(u)\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right) \mathrm{d}u \\ &= \int \mathcal{F}f(x) \left(\mathrm{K}_{h_m} * \mathbb{P}^Y\right)(x) \, \mathrm{d}x. \end{aligned} \tag{3.2.39}$$

Contrarily to the case of L^2-loss, the approximation error $|\theta - \theta_m|^2$ is no longer monotone in m and cannot, as previously, be simplified, since Pythagoras' theorem no longer applies.

We adopt here the strategy which has first been proposed in [44] and then used in [9] and consider an alternative criterion. Rather than the approximation error $|\theta - \theta_m|^2$, we consider the quantity $\sup_{k\geq m, k\in\mathcal{M}} |\theta_k - \theta_m|^2$, which can be estimated from the data.

For a detailed discussion on the underlying idea, we refer to [44].

Given the collection $\mathcal{M}$, we wish to choose the cutoff parameter $\widehat{m}$ in such way that for the corresponding risk, the following oracle inequality holds:

$$\begin{aligned} &\mathbb{E}\left[\left|\theta - \widehat{\theta}_{\widehat{m}}\right|^2\right] \\ \leq\ & C \inf_{m\in\mathcal{M}} \left\{ |\theta - \theta_{m_n}|^2 + \sup_{\substack{k\geq m\\ k\in\mathcal{M}}} |\theta_k - \theta_m|^2 + \mathrm{pen}(m) \right\} + C' n^{-1}, \end{aligned} \tag{3.2.40}$$

with constants C and C' which do not depend on the unknown smoothness parameters.

In the setting of density deconvolution with *known* distribution of the noise, and with K denoting the sinc kernel (see [9] for the details about this approach), the way to go is to replace the unknown quantities $|\theta_k - \theta_m|^2$ appearing in the oracle inequality by their corrected version $|\widehat{\theta}_k - \widehat{\theta}_m|^2 - \mathrm{H}^2(m,k)$. The deterministic bias correction term $\mathrm{H}^2(m,k)$ is then chosen to be

$$\mathrm{H}^2(m,k) := \frac{1}{n}\gamma\lambda_{m,k}\left(\sigma^2_{m,k} \vee x^2_{m,k}\right), \tag{3.2.41}$$

where $\lambda_{m,k}$ are logarithmic weights to be appropriately chosen, γ is some purely numerical constant, $\sigma^2_{m,k}$ is given by

$$\begin{aligned} \sigma^2_{m,k} \ :=\ & \frac{1}{4\pi^2}\left\{ \|\varphi_\varepsilon\|_{L^1} \int_{\{\pi m\leq|u|\leq\pi k\}} \left|\frac{\mathcal{F}f(-u)}{\varphi_\varepsilon(u)}\right|^2 \mathrm{d}u \right. \\ & \left. \wedge \left(\int_{\{\pi m\leq|u|\leq\pi k\}} \left|\frac{\mathcal{F}f(-u)}{\varphi_\varepsilon(u)}\right| \mathrm{d}u \right)^2 \right\} \end{aligned} \tag{3.2.42}$$

and

$$x_{m,k} := \frac{1}{\sqrt{n}}\frac{1}{2\pi} \int_{\{\pi m\leq|u|\leq\pi k\}} \left|\frac{\mathcal{F}f(-u)}{\varphi_\varepsilon(u)}\right| \mathrm{d}u. \tag{3.2.43}$$

The cutoff parameter $\widehat{m}$ is then chosen as the minimiser in $\mathcal{M}$ of the penalized criterion

$$\widehat{\mathrm{Crit}}(m) := \sup_{\substack{k\geq m\\ k\in\mathcal{M}}} \left\{ |\widehat{\theta}_k - \widehat{\theta}_m|^2 - \mathrm{H}^2(m,k) \right\} + \mathrm{pen}(m), \tag{3.2.44}$$

with $\mathrm{pen}(m) = \mathrm{H}(0, m)$.

To deal, in the present setting, with the unknown φ_ε, we could now try and proceed as in the preceding section and apply the same global thresholding scheme, thus replacing σ_m^2 as well as $\mathrm{H}^2(m, k)$ by their empirical versions, with $\frac{1}{\check{\varphi}_n}$ defined as in Definition 3.2.1.

However, this approach is unsatisfactory when we replace the sinc kernel by a general kernel function. The reason is that the argument given in (3.2.20) no longer applies in this situation. For this reason, we need a refined argument.

Still, there is some point in arguing that for large enough values of $|\varphi(u)|$, $\frac{1}{\widetilde{\varphi}_n(u)}$ should be be close to $\frac{1}{\varphi(u)}$, not only pointwise, but uniformly. But what happens for small values of $|\varphi(u)|$? In this case, $\left|\frac{1}{\widetilde{\varphi}_n(u)}\right|$ is certainly not a reasonable estimate, but a systematic underestimate of $\frac{1}{|\varphi(u)|}$.

Consequently, there is no hope that the stochastic correction term could be close to the true one and penalizing with $\widetilde{\mathrm{H}}^2(m, k)$ rather than with $\mathrm{H}^2(m, k)$ seems hopeless.

However, we might try some change of perspective: It is common, in a way, (see [18] and [20]) in model selection with unknown variance, to start working with some deterministic penalty term, which is, in fact, not feasible and see that the model selection procedure works. At the second stage, this deterministic (but unknown) quantity is replaced by a stochastic counterpart and one has to argue that the stochastic quantity is close to the deterministic one. This fails to hold for small values of $|\varphi(u)|$.

Still, when going one step back: What is the use in penalizing or introducing some bias correction at all? If the distribution of the noise is known, we start by considering some deterministic criterion

$$\mathrm{Crit}(m) = \sup_{\substack{k>m \\ k\in\mathcal{M}}} |\theta_k - \theta_m| + \mathrm{pen}(m). \tag{3.2.45}$$

Next, $|\theta_k - \theta_m|^2$ is replaced by its bias (over) corrected version $|\widehat{\theta}_k - \widehat{\theta}_m|^2 - \mathrm{H}^2(m, k)$.

However, if the distribution of the noise is unknown, it is certainly true that $\widetilde{\mathrm{H}}(m, k)^2$ is a systematic underestimate of $\mathrm{H}^2(m, k)$. But on the other hand, $\frac{1}{\overset{\approx}{\varphi}}$ is involved in the definition of $|\widehat{\theta}_m - \widehat{\theta}_k|^2$ as well as in the definition of $\widetilde{\mathrm{H}}^2(m, k)$ and if $\left|\frac{1}{\widetilde{\varphi}_n(u)}\right|$ is small compared to $\left|\frac{1}{\varphi(u)}\right|$, there is certainly no danger at all that $|\widehat{\theta}_k - \widehat{\theta}_m|^2$ could be (pointwise) an *over*estimate of $|\theta_k - \theta_m|^2$, so there is simply no need to subtract some large correction term.

We conclude that $\widetilde{\mathrm{H}}^2(m, k)$ being an underestimate simply does not matter since the phenomenon of underestimating $\frac{1}{\varphi}$ arises in the definition of $|\widehat{\theta}_k - \widehat{\theta}_m|$ as well.

So everything boils down to give some argument why $\frac{1}{\overset{\approx}{\varphi}}$ should, in an appropriate sense, be close to $\frac{1}{\varphi}$ not only pointwise nor on some compact set but *uniformly on the real line.*

To be able to do this, let us introduce another alternative estimator of $\frac{1}{\varphi}$.

Definition 3.2.5. *Let the weight function w be defined as in Definition 3.2.1. Let the truncated version of $\widehat{\varphi}_n(u)$ be*

$$\check{\varphi}_n(u) := \check{\varphi}_n^{\kappa,\delta} := \begin{cases} \widehat{\varphi}_n(u), & \text{if } |\widehat{\varphi}_n(u)| \geq \kappa(\log n)^{1/2} w(u)^{-1} n^{-1/2} \\ \kappa^{-1}(\log n)^{-1/2} w(u) n^{1/2}, & \text{else.} \end{cases} \tag{3.2.46}$$

Let the corresponding estimator of $\frac{1}{\varphi(u)}$ be

$$\frac{1}{\check{\varphi}_n(u)} = \frac{1}{\check{\varphi}_n^{\kappa,\delta}(u)}. \tag{3.2.47}$$

Introducing the extra logarithmic factor in the definition of $\frac{1}{\check{\varphi}}$ which will enable us to apply concentration inequalities of Talagrand type. This will be the key to proving the following result, which makes the original result by Neumann uniform on the real line:

Lemma 3.2.6. *Let c_1 be the constant appearing in Talagrand's inequality (see Lemma 3.3.3). Let $\frac{1}{\check{\varphi}_n}$ be defined by (3.2.47) with κ be chosen such that for some $\gamma > 0$, we have $\kappa \geq 2(\sqrt{2c_1} + \gamma)$. Then we have for some constant C_κ depending on the choice of κ, γ and δ that for $n \geq 2$:*

$$\mathbb{E}\left[\sup_{u \in \mathbb{R}} \frac{\left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2}{\frac{(\log n) w(u)^{-2} n^{-1}}{|\varphi(u)|^4} \wedge \frac{1}{|\varphi(u)|^2}}\right] \leq C_\kappa. \tag{3.2.48}$$

Indeed, the fact that we can show not only pointwise, but uniform closeness of $\frac{1}{\check{\varphi}_n}$ to $\frac{1}{\varphi}$ will be essential to make the model selection procedure work.

Now, we proceed as follows: The above definition gives rise to the following redefinition of $\widehat{\theta}_m$:

Definition 3.2.7.

(i) In the convolution model, let K *be some kernel functions for which (K1) and (K2) are satisfied. For $m \in \mathbb{N}$, let the kernel estimator corresponding to the bandwidth $\frac{1}{m}$ be defined to be*

$$\widehat{\theta}_m := \frac{1}{2\pi} \int \mathcal{F} f(-u) \frac{\widehat{\varphi}_{Z_n}}{\check{\varphi}_n(u)} \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right) du. \tag{3.2.49}$$

(ii) In the Lévy model, let

$$\widehat{\theta}_m := \frac{1}{2\pi} \int \mathcal{F} f(-u) \frac{\widehat{\varphi}_n'(u)}{i\check{\varphi}_n(u)} \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right) du. \tag{3.2.50}$$

Next, let us introduce the stochastic correction terms corresponding to $\frac{1}{\check{\varphi}_n}$:

Definition 3.2.8.

(i) *In the model of density deconvolution, let*

$$\widetilde{\mathrm{H}}^2(m,k) := n^{-1}\left\{c^{\mathrm{pen}} c_1 \widetilde{\lambda}^2_{m,k} + 16\left(\frac{5}{2}\kappa\right)^2 (\log n)\right\}\left(\widetilde{\sigma}^2_{m,k} \vee \widetilde{x}^2_{m,k}\right) \tag{3.2.51}$$

with

$$\widetilde{\sigma}^2_{m,k} = \frac{1}{2\pi^2}\Bigg\{\overline{C}_D \int \left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|^2 \left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|^2 w(u)^{-2}\,du$$
$$\wedge\ 3\left(\int \left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right| \left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right| w(u)^{-1}\,du\right)^2\Bigg\} \tag{3.2.52}$$

and

$$\widetilde{x}_{m,k} = \frac{1}{\sqrt{n}}\frac{1}{2\pi}\int \left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right| \left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right| w(u)^{-1}\,du \tag{3.2.53}$$

and

$$\widetilde{\lambda}_{m,k} := \log\left(\widetilde{\sigma}^2_{m,k}(k-m)^2\right) \vee 2\log\left(\widetilde{x}_{m,k}(k-m)\right). \tag{3.2.54}$$

For some $\gamma > 0$, *let* $c^{\mathrm{pen}} = 64 \vee 16(2c_1 + \gamma)$ *and and* $\kappa = 2(\sqrt{4c_1} + \gamma)$. *Moreover, let* $\overline{C}_D$ *be some positive constant.*

Let

$$\widetilde{\mathrm{pen}}(m) := \widetilde{\mathrm{H}}^2(0,m). \tag{3.2.55}$$

(ii) *In the Lévy model, let* $\widetilde{\mathrm{H}}^2(m,k)$ *be defined as in (3.2.51), and let* $\widetilde{x}_{m,k}$ *be defined as in (3.2.53).*

Let $\widetilde{\sigma}^2_{m,k}$ *be defined as in (3.2.52), apart from the fact that* $\overline{C}_D$ *is replaced by some constant* $\overline{C}_1 \leq \infty$ *and* $3C$ *is replaced by some constant* $\overline{C}_2$.

Let the weights be defined as follows: For some constant $\eta > 0$, *let*

$$\widetilde{\lambda}_{m,k} := \frac{8}{\eta}\log\left(\log(n\widetilde{x}_{m,k}(k-m))\right)^2 \log(n\widetilde{x}_{m,k}(k-m))\log\left(\widetilde{x}^2_{m,k}(k-m)^2\right)$$
$$\vee \log\left(\widetilde{\sigma}^2_{m,k}(k-m)^2\right). \tag{3.2.56}$$

We understand by $\mathrm{pen}(m)$ and $\mathrm{H}^2(m,k)$ the deterministic versions of $\widetilde{\mathrm{pen}}(m)$ and $\widetilde{\mathrm{H}}^2(m,k)$, that is, the definitions are the same as in formula (3.2.51) and formula (3.2.55), apart from the fact that $\frac{1}{\check{\varphi}_n}$ is replaced by $\frac{1}{\varphi}$.

These definitions give rise to the following choice of the cutoff parameter

Definition 3.2.9. *With the above definitions of* $\widehat{\theta}_m$, $\widetilde{\mathrm{H}}^2(m,k)$ *and* $\widetilde{\mathrm{pen}}(m)$, *be it in the Lévy model or in the convolution model, let*

$$\widehat{m} \ := \ \operatorname*{arginf}_{m\in\mathcal{M}}\ \widehat{\mathrm{Crit}}(m) \tag{3.2.57}$$

$$:= \operatorname*{arginf}_{m\in\mathcal{M}} \sup_{\substack{k>m\\k\in\mathcal{M}}} \left\{ |\widehat{\theta}_k - \widehat{\theta}_m|^2 - \widetilde{\mathrm{H}}^2(m,k) \right\} + \widetilde{\mathrm{pen}}(m).$$

Then we can formulate the following statements and hence the main results of this section:

Theorem 3.2.10. *Let observations according to the deconvolution model with unknown error density be given. Assume that (D1)-(D4) are satisfied. Assume, moreover, that (K1) and (K2) and (F1) or (F2) are satisfied. Assume that $C_D \leq \bar{C}_D$. Let a collection $\mathcal{M} \subseteq \mathbb{N}$ of indices be given and let $\widehat{m}$ be defined according to Definition 3.2.9 . Then we find for the risk of the corresponding estimator $\widehat{\theta}_{\widehat{m}}$ that*

$$\mathbb{E}\left[|\theta - \widehat{\theta}_{\widehat{m}}|^2\right] \leq C^{ad} \inf_{m\in\mathcal{M}} \left\{ |\theta - \theta_{m_n}|^2 + \sup_{\substack{k>m\\k\in\mathcal{M}}} |\theta_k - \theta_m|^2 + \mathrm{pen}(m) \right\} + C' n^{-1}$$

with constants C^{ad} and C' depending on the choice of γ and δ, but not on the unknown underlying parameters.

An analogous result can be stated for the Lévy model. However, we need here an exponential moment assumption.

Theorem 3.2.11. *Let observations $X_1, \cdots, X_{2n}$ according to the Lévy model be given. Assume that the assumptions (A1)- (A5) are satisfied. Assume, moreover, that (F1) or (F2) and (K1) or (K2) are met. Assume that $C_1 \leq \bar{C}_1$ and $C_2 \leq \bar{C}_2$. Assume, moreover, that $\mathbb{E}\left[\exp\left(\eta|X_1|\right)\right] < \infty$. Let $\widehat{m}$ be defined as in Definition 3.2.9. Then we can estimate*

$$\mathbb{E}\left[|\theta - \widehat{\theta}_{\widehat{m}}|^2\right] \leq C^{ad} \inf_{m\in\mathcal{M}} \left\{ |\theta - \theta_{m_n}|^2 + \sup_{\substack{k>m\\k\in\mathcal{M}}} |\theta_k - \theta_m|^2 + \mathrm{pen}(m) \right\} + C' n^{-1}$$

with some constant C^{ad} depending on the particular choices of the constants, but not on the unknown parameters and with a constant C' depending on the choice of the constants and on $\eta^{-1}\,\mathbb{E}\left[\exp\left(\eta|X_1|\right)\right]$.

The exponential moment assumption formulated in the preceding statement can be relaxed, but at the cost of losing a polynomial factor and limiting the size of the model and the choice of the kernel.

Theorem 3.2.12. *In the situation of Theorem 3.2.11, let $\bar{C}_1 = \infty$. Let $m_n = n$. Moreover, let f be such that $\|\mathcal{F}f\|_\infty < \infty$ and let $\mathcal{F}\,\mathrm{K}$ be supported in $[-\pi, \pi]$. Let $\bar{C}_2 \geq 1$ and assume that $C_2 \leq \bar{C}_2$. Assume, moreover, that $\mathbb{E}\left[|X_1|^{10}\right] < \infty$. Again, let $\widehat{m}$ be defined as in Definition 3.2.9. Then the risk of the corresponding estimator can be estimated as follows:*

$$\mathbb{E}\left[|\theta - \widehat{\theta}_{\widehat{m}}|^2\right] \leq C^{ad} \inf_{m\in\mathcal{M}} \left\{ |\theta - \theta_{m_n}|^2 + \sup_{\substack{k>m\\k\in\mathcal{M}}} |\theta_k - \theta_m|^2 + \mathrm{pen}(m) \right\} + C' n^{-1},$$

with C^{ad} depending on the choice of the constants and with C' depending on the choice of the constants and on $\mathbb{E}\left[|X_1|^{10}\right]$.

When considering the asymptotic behaviour, one can derive from the above theorems that there will occur a logarithmic loss, compared to the minimax rates of convergence.

It is well known that estimators which are adaptive minimax for estimating linear functionals do typically not exist. In a white noise framework, adaptive lower bounds have been proved by Lepski, see [46, 47]. In a density deconvolution model, minimax adaptivity has been treated by Butucea and Comte [9].

Consequently, it is not surprising that a logarithmic loss is also found in the Lévy model and it seems highly probable that the results obtained for density deconvolution carry over to the present setting, using similar, but highly tedious and technical arguments.

It is remarkable that the above results do no longer depend on any preliminary assumptions on the decay of the characteristic function. Even the quasi-monotonicity condition formulated in Assumption 3.2.2 can now be disposed of.

Illustration: Simulation example

We assume that μ has a square integrable Lebesgue density g which is continuous away from zero.

We consider the estimation of $g(y_k),\ k=1,2$ for the points $y_1 = 0.5$ and $y_2 = 2$. Moreover, we consider the estimation of $\int_I g(x)\,\mathrm{d}x$ for the interval $I = [-1.25, -0.25]$.

We work with the Gaussian kernel and apply the adaptive estimation procedure to $\Gamma(5,2)$ and $\Gamma(2.5,2)$-processes, to symmetric bilateral Gamma processes with parameters 2.5 and 2 ($\mathrm{b}\Gamma(2.5,2)$), and to compound Poisson processes with intensity 5 and standard normal distribution of the jumps ($\mathrm{cP}\mathcal{N}(5,0,1)$).

The collection of smoothing parameters is chosen to be $\mathcal{M}_n = \{1, \cdots, \sqrt{n}\}$. The processes are observed at time points $\Delta, \cdots, 2n\Delta$ with $\Delta = 1$ and with $n = 500, 1000, 10000$.

The performance of the adaptive estimator $\widehat{\theta}_{\widehat{m},n}$ is compared to the “estimator" $\widehat{\theta}_{m^\star,n}$ with oracle bandwidth:

$$\mathbb{E}\left[\left|\theta - \widehat{\theta}_{m^\star,n}\right|^2\right] = \inf_{m\in\mathcal{M}_n} \mathbb{E}\left[\left|\theta - \widehat{\theta}_{m,n}\right|^2\right] \tag{3.2.58}$$

We consider $k = 250$ independent iterations of the adaptive estimation procedure, as well as of the “estimator" with oracle choice of the bandwidth and compare the empirical risks.

We set $\kappa = 2$ and work with the penalty term

$$\widetilde{\mathrm{H}}^2(m,k) := \left(32\widetilde{\lambda}_{m,k} + 8\log n\right)\widetilde{\sigma}^2_{m,k}. \tag{3.2.59}$$

Again, the constants are, in praxis, chosen small, compared to the theoretical choice. Simulation experiments indicate that the performance of the procedure could be substantially improved, in the present case, by choosing the constants even smaller. However, from a theoretical perspective, choosing the constants smaller is hard to justify, since this would contradict the assumption that $C_1 \leq \bar{C}_1$ holds true.

The optimal theoretical choice of the constants remains an interesting open question.

We denote by $\widehat{r}_{or}$ the empirical risk with oracle choice of the bandwidth, multiplied by 100, and by $\widehat{r}_{ad}$ the empirical risk of the adaptive estimator, multiplied by 100. The values are summarised in the table below:

		y_1		y_2		I	
	n	$\widehat{r}_{or}$	$\widehat{r}_{ad}$	$\widehat{r}_{or}$	$\widehat{r}_{ad}$	$\widehat{r}_{or}$	$\widehat{r}_{ad}$
$\Gamma(5,2)$	500	13,18	62,21	2,21	6,50	11,98	24,55
	1000	8,09	53,75	1,63	6,14	7,10	24.14
	10000	1,28	28,03	0,62	4,88	0,42	23,48
$\Gamma(2.5,2)$	500	0,77	10,00	0,41	1,39	0,06	5,92
	1000	0,40	6,99	0,30	1,40	0,02	5,93
	10000	0.11	2.47	0,08	0,82	0.01	1.62
$\mathrm{b}\Gamma(2.5,2)$	500	16,98	70,33	0,05	0,68	2,03	27,13
	1000	11,17	67,92	0,03	0,68	0,99	18,35
	10000	1,44	39,67	0.01	0.67	0,19	8,10
$\mathrm{cP}\mathcal{N}(5,0,1)$	500	37,72	68,85	0,75	2,28	42,19	86,56
	1000	32,42	68,73	0,43	1,65	32,93	79,10
	10000	4,26	18,71	0,28	0,90	10,52	52,85

Discussion: The numerical results show that the estimation procedure adapts automatically to the local smoothness of the density, since the performance of the adaptive procedure is, just as the performance of the estimator with oracle bandwidth far better at the point $y_2 = 2$ than at $y_1 = 0.5$, near the discontinuity at zero.

The adaptive estimation procedure seems to fail completely, in comparison with the oracle estimator, for estimating the value of the integral for $\Gamma(5, 2)$-processes. This phenomenon seems to be a consequence of the fact that σ_m^2 is an upper bound for the variance, but the variance term may be considerably smaller. Numerical experiments indicate that this is precisely what happens in the present situation.

3.3 Proofs

We start by restating, for the reader's convenience, the concentration inequalities which will be essential for our reasoning.

Lemma 3.3.1 (Bernstein's Inequality). *Let $X_1, \cdots, X_n$ be complex valued i.i.d. random variables with $\operatorname{Var}(X_1) \leq v^2$ and suppose that $\|X_1\|_\infty \leq B$ for some $B < \infty$. Let $S_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])$. Then the following holds true for arbitrary $\kappa > 0$:*

$$\mathbb{P}\left(\left\{|S_n| \geq \kappa\right\}\right) \leq 4\exp\left(-n\frac{\kappa^2}{4v^2 + \frac{4\sqrt{2}}{3}\kappa B}\right) \tag{3.3.1}$$

This result is classical and dates back to the 1920s. A proof can be found, for example in [23].

The following integral version of the classical Bernstein inequality can be derived readily from Lemma 3.3.1:

Lemma 3.3.2. *In the situation of the preceding lemma, suppose that $\mathbb{E}[|S_n|] \leq \mathrm{H}$. Then we have*

$$\mathbb{E}\left[\left\{|S_n|^2 - \mathrm{H}^2\right\}_+\right] \leq 32\frac{v^2}{n}\exp\left(-n\frac{\mathrm{H}^2}{8v^2}\right) + 128\sqrt{2}\frac{B^2}{n^2}\exp\left(-n\frac{\mathrm{H}}{\frac{16\sqrt{2}}{3}B}\right). \tag{3.3.2}$$

Finally, we need the Talagrand inequality, which strengthens the classical Bernstein inequality to countable sets of random variables.

Lemma 3.3.3 (Talagrand's inequality). *Let I be some countable index set. For each $i \in I$, let $X_1^{(i)}, \cdots, X_n^{(i)}$ be centred i.i.d. complex valued random variables, defined on the same probability space, with $\|X_1^{(i)}\|_\infty \leq B$ for some $B < \infty$. Let $v^2 := \sup_{i\in I}\operatorname{Var}X_1^{(i)}$. Then for arbitrary $\varepsilon > 0$, there are positive constants c_1 and $c_2 = c_2(\varepsilon)$ depending only on ε such that for any $\kappa > 0$:*

$$\mathbb{P}\left(\left\{\sup_{i\in I}|S_n^{(i)}| \geq (1+\varepsilon)\,\mathbb{E}\left[\sup_{i\in I}|S_n^{(i)}|\right] + \kappa\right\}\right) \leq 2\exp\left(-n\left(\frac{\kappa^2}{c_1 v^2} \wedge \frac{\kappa}{c_2 B}\right)\right). \tag{3.3.3}$$

Lemma 3.3.3 is taken from Massart [52], see formula (5.50) on page 170. From the arguments given therein, we derive that for $\eta \in (0,1)$, we can take $c_1 = 4/(1-\eta)^2$ and $c_2(\varepsilon) = 4\sqrt{2}(1/3 + \varepsilon^{-1})/\eta$.

3.3.1 Proofs of the main results of Section 3.2.2

We start by treating the general case which has been formulated in Section 3.2.2. This will simplify our reasoning and avoid redundancy, since the results of Section 3.2.1 can basically be treated as special cases.

Preliminaries

In what follows, we formulate and prove a series of auxiliary results which will be essential for proving the main results of Section 3.2.

We start by proving an exponential bound for the deviation of the empirical characteristic function from the true one, uniformly on the real line.

Lemma 3.3.4. *Let $\tau > 0$ be given. Let δ be the constant appearing in the definition of the weight function w and let c_1 be the constant in Talagrand's inequality. Then, for arbitrary $\gamma > 0$, there is a positive constant $C_\kappa = C_\kappa^{\tau,\gamma,\delta}$ depending only on the choice of τ, γ and δ such that we have for $n \geq 1$:*

$$\mathbb{P}\left(\left\{\exists u \in \mathbb{R} : |\widehat{\varphi}_n(u) - \varphi(u)| \geq \tau(\log n)^{1/2} w(u)^{-1} n^{-1/2}\right\}\right) \leq C_\kappa n^{-\frac{(\tau-\gamma)^2}{c_1}}.$$

Proof. We prove the claim for the countable set of rational numbers. By continuity of the characteristic function and of w, it carries over to the whole range of real numbers.

By Theorem 4.1 in [56], we have for some positive constant C_{NR}:

$$\mathbb{E}\left[\sup_{u \in \mathbb{R}} |\widehat{\varphi}_n(u) - \varphi(u)| w(u)\right] \leq C_{\mathrm{NR}} n^{-1/2}. \tag{3.3.4}$$

Since moreover, we trivially have

$$\sup_{u \in \mathbb{R}} \mathrm{Var}[\widehat{\varphi}_1(u)] \leq 1 \quad \text{and} \quad \sup_{u \in \mathbb{R}} \|\widehat{\varphi}_1(u) w(u)\|_\infty \leq 1, \tag{3.3.5}$$

we can apply Talagrand's inequality. Setting

$$\kappa_n := \tau(\log n)^{1/2} n^{-1/2} - (1+\varepsilon) C_{\mathrm{NR}} n^{-1/2}, \tag{3.3.6}$$

for some $\varepsilon > 0$, we can estimate

$$\begin{aligned}
& \mathbb{P}\left(\left\{\exists q \in \mathbb{Q} : |\widehat{\varphi}_n(q) - \varphi(q)| \geq \tau(\log n)^{1/2} w(q)^{-1} n^{-1/2}\right\}\right) \\
= \ & \mathbb{P}\left(\left\{\sup_{q \in \mathbb{Q}} |\widehat{\varphi}_n(q) - \varphi(q)| w(q) \geq \tau(\log n)^{1/2} n^{-1/2}\right\}\right) \\
\leq \ & \mathbb{P}\left(\left\{\sup_{q \in \mathbb{Q}} |\widehat{\varphi}_n(q) - \varphi(q)| w(q) \geq (1+\varepsilon)\, \mathbb{E}\left[\sup_{q \in \mathbb{Q}} |\widehat{\varphi}_n(q) - \varphi(q)| w(q)\right] + \kappa_n\right\}\right) \\
\leq \ & 2\exp\left(-n\left(\frac{\kappa_n^2}{c_1} \wedge \frac{\kappa_n}{c_2}\right)\right).
\end{aligned} \tag{3.3.7}$$

By definition of κ_n, we have for C_κ large enough and arbitrary $n \geq 1$:

$$\begin{aligned}
& 2\exp\left(-n\left(\frac{\kappa_n^2}{c_1} \wedge \frac{\kappa_n}{c_2}\right)\right) \\
& \leq 2\exp\left(-\frac{\left(\tau(\log n)^{1/2} - (1+\varepsilon) C_{\mathrm{NR}}\right)^2}{c_1}\right) \vee 2\exp\left(-\frac{n^{1/2}\left(\tau(\log n)^{1/2} - (1+\varepsilon) C_{\mathrm{NR}}\right)}{c_2}\right)
\end{aligned}$$

$$\leq C \exp\left(-\frac{(\tau-\gamma)^2}{c_1}(\log n)\right) = Cn^{-\frac{(\tau-\gamma)^2}{c_1}}. \tag{3.3.8}$$

This is the desired result for the rational numbers and hence, by continuity, for the real line. □

We can now use Lemma 3.3.4 to analyse the deviation of $\frac{1}{\check{\varphi}_n}$ from $\frac{1}{\varphi}$.

Not surprisingly, the fact that we have introduced in the definition of $\frac{1}{\check{\varphi}_n}$ an extra logarithmic factor is essential for the analysis, since this is the key to making the pointwise result uniform on the real line. Of course, the price we have to pay is that we will be losing a logarithmic factor.

Proposition 3.3.5. *Let* $\frac{1}{\check{\varphi}_n} = \frac{1}{\check{\varphi}_n^{\kappa,\delta}}$ *be defined according to Definition 3.2.5. Assume that for some* $\gamma > 0$ *and some* $p > 0$*, we have* $\kappa \geq 2(\sqrt{pc_1} + \gamma)$*, where* c_1 *denotes the constant in Talagrand's inequality. Then we find that for* $n \geq 1$*,*

$$\mathbb{P}\left(\left\{\exists u \in \mathbb{R} : \left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2 > \left((4\kappa)^2 \frac{(\log n) w(u)^{-1} n^{-1}}{|\varphi(u)|^4} \wedge \left(\frac{5}{2}\right)^2 \frac{1}{|\varphi(u)|^2}\right)\right\}\right)$$
$$\leq Cn^{-p}.$$

Proof. Let us introduce the favourable set

$$C := C^{\kappa,\delta} := \left\{\forall u \in \mathbb{R} : |\widehat{\varphi}_n(u) - \varphi(u)| \leq \frac{\kappa}{2}(\log n)^{1/2} w(u)^{-1} n^{-1/2}\right\}. \tag{3.3.9}$$

We start by recalling that, thanks to Lemma 3.3.4 and to the choice of the constant κ, we have,

$$\mathbb{P}(C^c) \leq Cn^{-\frac{(\kappa/2-\gamma)^2}{c_1}} \leq C(n^{-p}), \tag{3.3.10}$$

so it is enough to consider the set C.

Now, let us introduce the following partition of the real line: We have $\mathbb{R} = \mathbb{R}_1^\kappa \cup \mathbb{R}_2^\kappa \cup \mathbb{R}_3^\kappa$, with

$$\mathbb{R}_1^\kappa = \left\{u \in \mathbb{R} : |\varphi(u)| < \frac{\kappa}{2}(\log n)^{1/2} w(u)^{-1} n^{-1/2}\right\}, \tag{3.3.11}$$

$$\mathbb{R}_2^\kappa = \left\{u \in \mathbb{R} : |\varphi(u)| > \frac{3}{2}\kappa(\log n)^{1/2} w(u)^{-1} n^{-1/2}\right\} \tag{3.3.12}$$

and

$$\mathbb{R}_3^\kappa = \left\{u \in \mathbb{R} : \frac{\kappa}{2} w(u)^{-1}\left(\frac{\log n}{n}\right)^{1/2} \leq |\varphi(u)| \leq \frac{3}{2}\kappa w(u)^{-1}\left(\frac{\log n}{n}\right)^{1/2}\right\}. \tag{3.3.13}$$

Consider first the set $\mathbb{R}_1^\kappa$. By definition of C, we find that for arbitrary $u \in \mathbb{R}_1^\kappa$, we have

$$|\widehat{\varphi}_n(u)| \leq |\varphi(u)| + |\varphi(u) - \widehat{\varphi}_n(u)| < \kappa(\log n)^{1/2} w(u)^{-1} n^{-1/2} \tag{3.3.14}$$

and hence by definition of $\frac{1}{\check{\varphi}_n}$:

$$\begin{aligned}\left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2 &= \frac{|\varphi(u) - \check{\varphi}_n(u)|^2}{|\varphi(u)|^2|\check{\varphi}_n(u)|^2} \\ &= \frac{|\varphi(u) - \kappa(\log n)^{1/2}w(u)^{-1}n^{-1/2}|^2}{|\check{\varphi}_n(u)|^2|\varphi(u)|^2} \\ &\leq \frac{\left(\frac{3}{2}\kappa\right)^2(\log n)w(u)^{-2}n^{-1}}{|\varphi(u)|^2|\check{\varphi}_n(u)|^2}. \qquad (3.3.15)\end{aligned}$$

Next, notice that, using again the definition of C and $\mathbb{R}_1^\kappa$,

$$\begin{aligned}\frac{1}{|\check{\varphi}_n(u)|^2} &= \kappa^{-2}(\log n)^{-1}w(u)^2 n \\ &= \frac{1}{4}\left(\frac{\kappa}{2}\right)^{-2}(\log n)^{-1}w(u)^2 n \leq \frac{1}{4}\frac{1}{|\varphi(u)|^2}. \qquad (3.3.16)\end{aligned}$$

Putting (3.3.15) and (3.3.16) together, we have shown that on C, we have for any $u \in \mathbb{R}_1^\kappa$:

$$\begin{aligned}&\left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2 \\ &\leq \left(\frac{3}{4}\kappa\right)^2\frac{(\log n)w(u)^{-2}n^{-1}}{|\varphi(u)|^4} \wedge \frac{9}{4}\frac{1}{|\varphi(u)|^2}. \qquad (3.3.17)\end{aligned}$$

Next, consider the set $\mathbb{R}_2^\kappa$. By definition of C, we find that on this set, for arbitrary $u \in \mathbb{R}_2^\kappa$

$$|\widehat{\varphi}_n(u)| \geq |\varphi(u)| - |\varphi(u) - \widehat{\varphi}_n(u)| \geq \kappa(\log n)^{1/2}w(u)^{-1}n^{-1/2} \qquad (3.3.18)$$

and hence $\check{\varphi}_n(u) = \widehat{\varphi}_n(u)$. This implies

$$\begin{aligned}\left|\frac{1}{\check{\varphi}_n} - \frac{1}{\varphi(u)}\right|^2 &= \frac{|\check{\varphi}_n(u) - \varphi(u)|^2}{|\check{\varphi}_n(u)|^2|\varphi(u)|^2} \\ &= \frac{|\widehat{\varphi}_n(u) - \varphi(u)|^2}{|\check{\varphi}_n(u)|^2|\varphi(u)|^2} \\ &\leq \frac{\left(\frac{\kappa}{2}\right)^2(\log n)w(u)^{-2}n^{-1}}{|\check{\varphi}_n(u)|^2|\varphi(u)|^2}. \qquad (3.3.19)\end{aligned}$$

By definition of $\frac{1}{\check{\varphi}_n}$, we always have $\frac{1}{|\check{\varphi}_n(u)|^2} \leq \kappa^{-2}(\log n)^{-1}w(u)n$, so we find that

$$\frac{\left(\frac{\kappa}{2}\right)^2(\log n)w(u)^{-2}n^{-1}}{|\check{\varphi}_n(u)|^2|\varphi(u)|^2} \leq \frac{1}{4}\frac{1}{|\varphi(u)|^2}. \qquad (3.3.20)$$

On the other hand, by definition $\mathbb{R}_2^\kappa$, we find on C:

$$|\widehat{\varphi}_n(u)| \quad \geq \quad |\varphi(u)| - |\widehat{\varphi}_n(u) - \varphi(u)|$$

$$\begin{aligned}
&\geq |\varphi(u)| - \frac{1}{2}\kappa(\log n)^{1/2} w(u)^{-1} n^{-1/2} \\
&\geq |\varphi(u)| - \frac{1}{3}|\varphi(u)| \\
&= \frac{2}{3}|\varphi(u)|. \qquad (3.3.21)
\end{aligned}$$

This allows to continue from (3.3.19):

$$\begin{aligned}
\frac{\left(\frac{\kappa}{2}\right)^2 (\log n) w(u)^{-2} n^{-1}}{|\check{\varphi}_n(u)|^2 |\varphi(u)|^2} &= \frac{\left(\frac{\kappa}{2}\right)^2 (\log n) w(u)^{-2} n^{-1}}{|\widehat{\varphi}_n(u)|^2 |\varphi(u)|^2} \\
&\leq \frac{\left(\frac{3}{4}\kappa\right)^2 (\log n) w(u)^{-2} n^{-1}}{|\varphi(u)|^4}. \qquad (3.3.22)
\end{aligned}$$

It remains to consider $\mathbb{R}_3^\kappa$. The triangle inequality gives on C

$$\begin{aligned}
&\left| \frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right|^2 \\
= \;& \frac{|\kappa(\log n)^{1/2} w(u)^{-1} n^{-1/2} - \varphi(u)|^2}{|\check{\varphi}_n(u)|^2 |\varphi(u)|^2} 1\left(\left\{ |\widehat{\varphi}_n(u)| < \kappa \left(\frac{\log n}{n}\right)^{1/2} w(u) \right\} \right) \\
+ \;& \frac{|\widehat{\varphi}_n(u) - \varphi(u)|^2}{|\check{\varphi}_n(u)|^2 |\varphi(u)|^2} 1\left(\{ |\widehat{\varphi}_n(u)| \geq \kappa(\log n)^{1/2} w(u) n^{-1/2} \} \right) \\
\leq \;& \frac{\left(\frac{5}{2}\kappa\right)^2 (\log n) w(u)^{-2} n^{-1}}{|\check{\varphi}_n(u)|^2 |\varphi(u)|^2}. \qquad (3.3.23)
\end{aligned}$$

Next, observe that on $\mathbb{R}_3^\kappa$, we have

$$\begin{aligned}
\left| \frac{1}{\check{\varphi}_n(u)} \right| &\leq \frac{3}{2}\left(\frac{3}{2}\kappa\right)^{-1} (\log n)^{-1/2} w(u) n^{1/2} \\
&\leq \frac{3}{2}\frac{1}{|\varphi(u)|}. \qquad (3.3.24)
\end{aligned}$$

From this we conclude that

$$\begin{aligned}
&\frac{\left(\frac{5}{2}\kappa\right)^2 (\log n) w(u)^{-2} n^{-1}}{|\check{\varphi}_n(u)|^2 ||\varphi(u)|^2} \\
\leq \;& \left(\frac{15}{4}\kappa\right)^2 \frac{(\log n) w(u)^{-2} n^{-1}}{|\varphi(u)|^4} \wedge \left(\frac{5}{2}\right)^2 \frac{1}{|\varphi(u)|^2}. \qquad (3.3.25)
\end{aligned}$$

Putting the above results together, we have shown the desired result on the favourable set C and hence, since C^c is negligible, the statement of the proposition. □

Note that we have shown the following important corollary, which is an immediate consequence of the proof of the preceding theorem:

Corollary 3.3.6. *In the situation of the preceding statement, we have*

$$\mathbb{P}\left(\left\{\exists u\in\mathbb{R}:\left|\frac{1}{\check{\varphi}_n(u)}-\frac{1}{\varphi(u)}\right|^2>\left(\frac{5}{2}\kappa\right)^2\frac{(\log n)w(u)^{-2}n^{-1}}{|\check{\varphi}_n(u)|^2|\varphi(u)|^2}\right\}\right)\le Cn^{-p}.$$

Proof. This is a direct consequence of the proof of Proposition 3.3.5. The statement of the corollary can be found in formulae (3.3.15), (3.3.19) and (3.3.23). □

It is in fact this version of the statement which will play an important role in the sequel.

The uniform version of the classical Neumann Lemma can now be stated as an easy consequence of Proposition 3.3.5.

Proof of Lemma 3.2.6. Let the set C be defined as in the proof of Proposition 3.3.5. We can decompose

$$\begin{aligned}
&\mathbb{E}\left[\sup_{u\in\mathbb{R}}\frac{\left|\frac{1}{\check{\varphi}_n(u)}-\frac{1}{\varphi(u)}\right|^2}{\frac{(\log n)w(u)^{-2}n^{-1}}{|\varphi(u)|^4}\wedge\frac{1}{|\varphi(u)|^2}}\right]\\
=\;&\mathbb{E}\left[\sup_{u\in\mathbb{R}}\frac{\left|\frac{1}{\check{\varphi}_n(u)}-\frac{1}{\varphi(u)}\right|^2}{\frac{(\log n)w(u)^{-2}n^{-1}}{|\varphi(u)|^4}\wedge\frac{1}{|\varphi(u)|^2}}1(C)\right]\\
+\;&\mathbb{E}\left[\sup_{u\in\mathbb{R}}\frac{\left|\frac{1}{\check{\varphi}_n(u)}-\frac{1}{\varphi(u)}\right|^2}{\frac{(\log n)w(u)^{-2}n^{-1}}{|\varphi(u)|^4}\wedge\frac{1}{|\varphi(u)|^2}}1(C^c)\right]
\end{aligned}\tag{3.3.26}$$

The definition of C, together with Proposition 3.3.5, readily implies that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{u\in\mathbb{R}}\frac{\left|\frac{1}{\check{\varphi}_n(u)}-\frac{1}{\varphi(u)}\right|^2}{\frac{(\log n)w(u)^{-2}n^{-1}}{|\varphi(u)|^4}\wedge\frac{1}{|\varphi(u)|^2}}1(C)\right]\\
\le\;&\mathbb{E}\left[\sup_{u\in\mathbb{R}}\frac{(4\kappa)^2\frac{(\log n)w(u)^{-1}n^{-1}}{|\varphi(u)|^4}\wedge\left(\frac{5}{2}\right)^2\frac{1}{|\varphi(u)|^2}}{\frac{(\log n)w(u)^{-2}n^{-1}}{|\varphi(u)|^4}\wedge\frac{1}{|\varphi(u)|^2}}1(C)\right]\le 16\kappa^2.
\end{aligned}\tag{3.3.27}$$

On the other hand, since we have, by definition,

$$\frac{1}{|\check{\varphi}_n(u)|}\le\kappa^{-1}(\log n)^{-1/2}w(u)n^{1/2},\tag{3.3.28}$$

we can estimate for arbitrary $u\in\mathbb{R}$:

$$\begin{aligned}
\frac{\left|\frac{1}{\check{\varphi}_n(u)}-\frac{1}{\varphi(u)}\right|^2}{\frac{(\log n)w(u)^{-2}n^{-1}}{|\varphi(u)|^4}\wedge\frac{1}{|\varphi(u)|^2}}&\le\frac{\left(\kappa^{-1}(\log n)^{-1/2}w(u)n^{1/2}+\frac{1}{|\varphi(u)|}\right)^2}{\frac{(\log n)w(u)^{2}n^{-1}}{|\varphi(u)|^4}\wedge\frac{1}{|\varphi(u)|^2}}\\
&\le(\kappa^{-1}+1)^2n^2
\end{aligned}\tag{3.3.29}$$

which yields

$$\mathbb{E}\left[\sup_{u\in\mathbb{R}} \frac{\left|\frac{1}{\tilde{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2}{\frac{(\log n)w(u)^{-2}n^{-1}}{|\varphi(u)|^4} \wedge \frac{1}{|\varphi(u)|^2}} \mathbb{1}(C^c)\right] \leq (\kappa^{-1}+1)^2 n^2\, \mathbb{P}\left(C^c\right). \tag{3.3.30}$$

Since we have shown, in Lemma 3.3.4 that

$$\mathbb{P}(C^c) \leq C_\kappa n^{-\frac{(\kappa/2-\gamma)^2}{c_1}} \tag{3.3.31}$$

holds true, and we have $\kappa \geq 2(\sqrt{2c_1}+\gamma)$, this gives the desired result. □

The result immediately extends to powers different from 2. The following Corollary can be obtained, replacing in each step 2 by $2q$:

Corollary 3.3.7. *In the situation of the preceding statement, let $\kappa \geq 2\left(\sqrt{2qc_1}+\gamma\right)$ for some $q \in \mathbb{R}^+$. Then we have for some constant $C_\kappa = C_{\kappa q}^{\gamma,\delta,\varepsilon}$ depending on γ, q, δ and ε:*

$$\mathbb{E}\left[\sup_{u\in\mathbb{R}} \frac{\left|\frac{1}{\tilde{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^{2q}}{\frac{(\log n)^q w(u)^{-2q}n^{-q}}{|\varphi(u)|^{4q}} \wedge \frac{1}{|\varphi(u)|^{2q}}}\right] \leq C_\kappa. \tag{3.3.32}$$

Auxiliary results

So far, our considerations have been fairly general, without any reference to Lévy processes nor density deconvolution or to the particular model selection problems.

In the present section, we prove some auxiliary results which are the key to proving the main results of Section 3.2.

First, we formulate an extension of Lemma 3.3.4 and of Proposition 3.3.5.

In what follows, to avoid redundancy, we will use the following notation: When talking about the model of density deconvolution, we let

$$C_1 := C_D := C(\|\varphi_Y\|_{\mathrm{L}^1} + 2\|\varphi_Y\|_{L^2}^2) \tag{3.3.33}$$

and

$$C_2 := 3C \tag{3.3.34}$$

with C denoting the constant depending on Neumann's Lemma. In case that we are in the Lévy model, we have, as previously $C_1 := C\left(2\|\Psi'\|_{\mathrm{L}^2}^2 + \|\Psi''\|_{\mathrm{L}^1}\right)$ and $C_2 := C\left(2\|\Psi'\|_\infty^2 + \|\Psi''\|_\infty\right)$, which is the original definition.

At the moment, it makes no difference weather we consider density deconvolution or the Lévy model.

Lemma 3.3.8. *Let*

$$\begin{aligned} x^2_{f_{m,k}} &:= \frac{1}{2\pi^2}\Bigg\{ C_1 \int |\mathcal{F}f(-u)|^2 \left| \mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|^2 du \\ &\qquad \wedge C_2 \left(\int |\mathcal{F}f(-u)| \left| \mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right| du \right)^2 \Bigg\}. \end{aligned} \tag{3.3.35}$$

Moreover, let

$$\lambda_{f_{m,k}} := \sqrt{c_1} \log\left(x^2_{f_{m,k}} (k-m)^2 \right). \tag{3.3.36}$$

For some $\gamma > 0$, let $\kappa = 2(\sqrt{2pc_1} + \gamma)$. Then we have for some constant C_{K} depending on γ, δ and ε:

$$\begin{aligned} &\mathbb{P}\left(\left\{ \exists u \in \mathbb{R} : |\widehat{\varphi}_n(u) - \varphi(u)| \geq \left(\frac{\kappa}{2} (\log n)^{1/2} + \lambda_{f_{m,k}} \right) w(u)^{-1} n^{-1/2} \right\} \right) \\ \leq\ & C_{\mathrm{K}} n^{-p} x^{-2}_{f_{m,k}} (k-m)^{-2}. \end{aligned}$$

Proof. The proof runs along the same lines as the proof of Lemma 3.3.4, setting, this time

$$\kappa_n := \left(\frac{\kappa}{2} (\log n)^{1/2} + \lambda_{f_{m,k}} \right) n^{-1/2} - C_{\mathrm{NR}} n^{-1/2}.$$

Using again continuity of the (empirical) characteristic function, the Talagrand inequality and the choice of κ, we derive that for C_{K} chosen large enough,

$$\begin{aligned} &\mathbb{P}\left(\left\{ \exists u \in \mathbb{R} : |\widehat{\varphi}_n(u) - \varphi(u)| \geq \left(\frac{\kappa}{2} (\log n)^{1/2} + \lambda_{f_{m,k}} \right) w(u)^{-1} n^{-1/2} \right\} \right) \\ \leq\ & 2\exp\left(- \frac{\left(\frac{\kappa}{2} (\log n)^{1/2} + \lambda_{f_{m,k}} - C_{\mathrm{NR}} \right)^2}{c_1} \right) \\ & \vee 2\exp\left(- \frac{n^{1/2} \left(\frac{\kappa}{2} (\log n)^{1/2} + \lambda_{f_{m,k}} - C_{\mathrm{NR}} \right)}{c_2} \right) \\ \leq\ & C_{\mathrm{K}} \exp\left(- \frac{(\kappa/2 - \gamma)^2}{c_1} (\log n) - \log\left(x^2_{f_{m,k}} (k-m)^2 \right) \right) \\ =\ & C_{\mathrm{K}} n^{-p} x^{-2}_{f_{m,k}} (k-m)^{-2}. \end{aligned} \tag{3.3.37}$$

□

The above result implies the following extension of Corollary 3.3.5:

Corollary 3.3.9. *In the situation of the preceding statement, we have for some constant C_{K} depending on γ and δ:*

$$\begin{aligned} &\mathbb{P}\left(\left\{ \exists u \in \mathbb{R} : \left| \frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right|^2 > \frac{\left(\frac{5}{2}\kappa (\log n) + \lambda_{f_{m,k}} \right)^2 w(u)^{-2} n^{-1}}{|\check{\varphi}_n(u)|^2 |\varphi(u)|^2} \right\} \right) \\ \leq\ & C_{\mathrm{K}} n^{-p} x^{-2}_{f_{m,k}} (k-m)^{-2}. \end{aligned}$$

Proof. This statement is derived from Lemma 3.3.8 in the same way as Corollary 3.3.6 is derived from Lemma 3.3.4 and from the proof of Proposition 3.3.5. We only have to replace, in each step, $\frac{\kappa}{2}(\log n)^{\frac{1}{2}}/w(u)$ by $\left(\frac{\kappa}{2}(\log n)^{\frac{1}{2}} + \lambda_{m,k}\right)/w(u)$. $\square$

We are now ready to prove the auxiliary result which is most important for the proof of the main results.

To clarify the definitions appearing in part (i), part (ii) and part (iii), let us mention that the first part treats the case of estimating linear functionals in the density deconvolution model. The second part treats the Lévy model under the exponential moment assumption and the third part the Lévy model under the assumption that the moments up to order 10 are finite. All three statements are formulated for general kernel functions.

Proposition 3.3.10.

(i) *For $m \in \mathbb{N}$, let the kernel estimator $\widehat{\theta}_m$ be defined by (3.2.49). Let θ_m be defined by (3.2.39). Assume that the conditions which are summarised in Theorem 3.2.10 are satisfied. Then we can estimate for arbitrary $m \in \mathbb{N}$:*

$$\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}}\left\{\left|(\widehat{\theta}_k-\widehat{\theta}_m)-(\theta_k-\theta_m)\right|^2-\frac{1}{2}\widetilde{\mathrm{H}}(m,k)^2\right\}_+\right]\le Cn^{-1}$$

with some positive constant C depending only on the choice of the constants.

(ii) *Let $\widehat{\theta}_m$ be defined by (3.2.50). Let θ_m be defined by (3.2.38). Assume that the conditions which are summarised in Theorem 3.2.11 are satisfied. Then we can estimate for arbitrary $m \in \mathbb{N}$:*

$$\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}}\left\{\left|(\widehat{\theta}_k-\widehat{\theta}_m)-(\theta_k-\theta_m)\right|^2-\frac{1}{2}\widetilde{\mathrm{H}}(m,k)^2\right\}_+\right]\le Cn^{-1},$$

where C is a positive constant depending on the exponential moment.

(iii) *In the situation of part (ii), assume that the conditions which are summarised in Theorem 3.2.12 are satisfied. Then we can estimate for arbitrary $m \in \mathcal{M}$:*

$$\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathcal{M}}}\left\{\left|(\widehat{\theta}_k-\widehat{\theta}_m)-(\theta_k-\theta_m)\right|^2-\frac{1}{2}\widetilde{\mathrm{H}}(m,k)^2\right\}_+\right]\le Cn^{-1},$$

where C is a positive constant depending on the 10-th moment.

Proof.

(i) For arbitrary $m \in \mathbb{N}$, we write

$$\widetilde{\theta}_m := \frac{1}{2\pi} \int \mathcal{F}f(-u) \frac{\varphi_Z(u)}{\check{\varphi}_n(u)} \mathcal{F}\,\mathrm{K}\left(\frac{u}{m}\right) \mathrm{d}u. \tag{3.3.38}$$

We use the estimate

$$\begin{aligned} &\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}} \left\{\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\theta_k - \theta_m)\right|^2 - \frac{1}{2}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\right] \\ \le\;& 2\,\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}} \left\{\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\right] \\ +\;& 2\,\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}} \left\{\left|(\widetilde{\theta}_k - \widetilde{\theta}_m) - (\theta_k - \theta_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\right]. \end{aligned} \tag{3.3.39}$$

Consider first the expression appearing in the second line of formula (3.3.39). We can estimate, conditioning on $\widehat{\varphi}_n$:

$$\begin{aligned} &\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}} \left\{\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\right] \\ =\;& \mathbb{E}\left[\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}} \left\{\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ \middle| \widehat{\varphi}_n\right]\right] \\ \le\;& \mathbb{E}\left[\sum_{\substack{k>m\\k\in\mathbb{N}}} \mathbb{E}\left[\left\{\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ \middle| \widehat{\varphi}_n\right]\right]. \end{aligned} \tag{3.3.40}$$

Since $\widehat{\varphi}_n$ and $\widehat{\varphi}_{Z_n}$ are independent by construction, we have that, conditional on $\widehat{\varphi}_n$,

$$\begin{aligned} &(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m) \\ =\;& \frac{1}{n}\sum_{j=1}^{n} \frac{1}{2\pi} \int \mathcal{F}f(-u) \frac{\left(e^{iuZ_j} - \mathbb{E}\left[e^{iuZ_1}\right]\right)}{\check{\varphi}_n(u)} \left(\mathcal{F}\,\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\,\mathrm{K}\left(\frac{u}{m}\right)\right) \mathrm{d}u \end{aligned} \tag{3.3.41}$$

is the sum of independent, centred random variables.

For the (conditional) variance, we have

$$\mathbb{E}\left[\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m)\right|^2 \middle| \widehat{\varphi}_n\right]$$

$$
\begin{aligned}
&\leq \frac{1}{n}\frac{1}{(2\pi)^2}\Bigg\{2\|\varphi_Y\|^2_{\mathrm{L}^2}\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|^2\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|^2\mathrm{d}u \\
&\qquad\wedge 2\left(\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|\mathrm{d}u\right)^2\Bigg\} \\
&\leq \frac{1}{n}\widetilde{\sigma}^2_{m,k}\quad a.s. \qquad (3.3.42)
\end{aligned}
$$

and moreover, we find that, conditional on $\widehat{\varphi}_n$,

$$
\begin{aligned}
&\left\|\frac{1}{2\pi}\int\mathcal{F}f(-u)\frac{e^{iuZ_1}}{\check{\varphi}_n(u)}\left(\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right)\mathrm{d}u\right\|_\infty \\
&\leq \frac{1}{2\pi}\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|\mathrm{d}u \\
&\leq \sqrt{n}\,\widetilde{x}_{m,k}\quad a.s. \qquad (3.3.43)
\end{aligned}
$$

We can thus apply Lemma 3.3.2 to the conditional expectation and find that

$$
\begin{aligned}
&\mathbb{E}\left[\left\{\left|(\widehat{\theta}_k-\widehat{\theta}_m)-(\widetilde{\theta}_k-\widetilde{\theta}_m)\right|^2-\frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\middle|\widehat{\varphi}_n\right] \\
&\leq 32\frac{\widetilde{\sigma}^2_{m,k}}{n}\exp\left(-n\frac{\widetilde{\mathrm{H}}^2(m,k)}{64\widetilde{\sigma}^2_{m,k}}\right) \\
&+ 128\sqrt{2}\frac{\widetilde{x}^2_{m,k}}{n}\exp\left(-n\frac{\widetilde{\mathrm{H}}(m,k)}{\frac{8\times16\sqrt{2}}{3}\sqrt{n}\widetilde{x}_{m,k}}\right) \\
&\leq 32\frac{\widetilde{\sigma}^2_{m,k}}{n}\exp\left(-\frac{c^{\mathrm{pen}}\widetilde{\lambda}^2_{m,k}\left(\widetilde{\sigma}^2_{m,k}\vee\widetilde{x}^2_{m,k}\right)}{64\widetilde{\sigma}^2_{m,k}}\right) \\
&+ 128\sqrt{2}\frac{\widetilde{x}^2_{m,k}}{n}\exp\left(-\frac{c^{\mathrm{pen}}\widetilde{\lambda}_{m,k}\left(\widetilde{\sigma}_{m,k}\vee\widetilde{x}_{m,k}\right)}{\frac{8\times16\sqrt{2}}{3}\widetilde{x}_{m,k}}\right). \qquad (3.3.44)
\end{aligned}
$$

Since we have chosen $c^{pen}\geq 64$ and $\widetilde{\lambda}^2_{m,k}\geq\log\left(\widetilde{\sigma}^2_{m,k}(k-m)^2\right)$ as well as $\widetilde{\lambda}_{m,k}\geq\log\left(\widetilde{x}^2_{m,k}(k-m)^2\right)$ (see the definition of the weights given in formula (3.2.54)), we can continue by estimating

$$
\begin{aligned}
&32\frac{\widetilde{\sigma}^2_{m,k}}{n}\exp\left(-\frac{c^{\mathrm{pen}}\widetilde{\lambda}^2_{m,k}\left(\widetilde{\sigma}^2_{m,k}\vee\widetilde{x}^2_{m,k}\right)}{64\widetilde{\sigma}^2_{m,k}}\right) \\
&+ 128\sqrt{2}\frac{\widetilde{x}^2_{m,k}}{n}\exp\left(-\frac{c^{\mathrm{pen}}\widetilde{\lambda}_{m,k}\left(\widetilde{\sigma}_{m,k}\vee\widetilde{x}_{m,k}\right)}{\frac{8\times16\sqrt{2}}{3}\widetilde{x}_{m,k}}\right) \\
&\leq 32\frac{\widetilde{\sigma}^2_{m,k}}{n}\exp\left(-\widetilde{\lambda}^2_{m,k}\right)+128\sqrt{2}\frac{\widetilde{x}^2_{m,k}}{n}\exp\left(-\widetilde{\lambda}_{m,k}\right)
\end{aligned}
$$

$$\begin{aligned}
&\leq \quad 32\frac{\widetilde{\sigma}^2_{m,k}}{n}\widetilde{\sigma}^{-2}_{m,k}(k-m)^{-2} + 128\sqrt{2}\frac{\widetilde{x}^2_{m,k}}{n}\widetilde{x}^{-2}_{m,k}(k-m)^{-2} \\
&\leq \quad (32+128\sqrt{2})\frac{(k-m)^{-2}}{n} \quad a.s. \qquad (3.3.45)
\end{aligned}$$

We have thus shown that

$$\begin{aligned}
&\sum_{\substack{k>m\\k\in\mathbb{N}}} \mathbb{E}\left[\left\{\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ \middle| \widehat{\varphi}_n\right] \\
\leq \quad & \frac{32+128\sqrt{2}}{n}\sum_{k>m}(k-m)^{-2} \; a.s. \qquad (3.3.46)
\end{aligned}$$

And hence for some universal positive constant C':

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}}\left\{\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\right] \\
\leq \quad & \mathbb{E}\left[\sum_{\substack{k>m\\k\in\mathbb{N}}}\mathbb{E}\left[\left\{\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ \middle| \widehat{\varphi}_n\right]\right] \\
= \quad & C'n^{-1}. \qquad (3.3.47)
\end{aligned}$$

It remains to consider the expression appearing in the last line of formula (3.3.39).

Let us introduce, for arbitrary $m \leq k$, the favourable set

$$\begin{aligned}
&C(m,k) \\
:= &\left\{\forall u \in \mathbb{R} : \left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2 \leq \frac{\left(\left(\frac{5}{2}\kappa\right)(\log n)^{1/2} + \lambda_{f_{m,k}}\right)^2 w(u)^{-1}}{n|\check{\varphi}_n(u)|^2||\varphi(u)|^2}\right\}, \qquad (3.3.48)
\end{aligned}$$

with $\lambda_{f_{m,k}}$ defined as in Lemma 3.3.8. We can estimate

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}}\left\{\left|\left(\widetilde{\theta}_k - \widetilde{\theta}_m\right) - (\theta_k - \theta_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\right] \\
&\leq \mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}}\left\{\left|\left(\widetilde{\theta}_k - \widetilde{\theta}_m\right) - (\theta_k - \theta_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ 1\Big(C(m,k)\Big)\right] \\
&+ \mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}}\left\{\left|\left(\widetilde{\theta}_k - \widetilde{\theta}_m\right) - (\theta_k - \theta_m)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ 1\Big(C(m,k)^c\Big)\right]. \qquad (3.3.49)
\end{aligned}$$

The Cauchy Schwarz inequality and the fact that $|\varphi_Y(u)| \leq 1$ imply

$$
\begin{aligned}
&\left|\left(\widetilde{\theta}_k - \widetilde{\theta}_m\right) - (\theta_k - \theta_m)\right|^2 \\
&= \left|\frac{1}{2\pi}\int \mathcal{F}f(-u)\varphi_Z(u)\left(\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right)\left(\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right)\mathrm{d}u\right|^2 \\
&= \left|\frac{1}{2\pi}\int \mathcal{F}f(-u)\varphi_Y(u)\left(\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right)\varphi(u)\left(\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right)\mathrm{d}u\right|^2 \\
&\leq \frac{1}{(2\pi)^2}\Bigg\{\left(\int |\mathcal{F}f(-u)|\left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right||\varphi(u)|\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|\mathrm{d}u\right)^2 \\
&\quad \wedge \Bigg(\int |\varphi_Y(x)|^2\,\mathrm{d}x \\
&\qquad \int |\mathcal{F}f(-u)|^2\left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2|\varphi(u)|^2\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|^2\mathrm{d}u\Bigg)\Bigg\}.
\end{aligned}
\tag{3.3.50}
$$

Now, the definition of the set $C(m,k)$ readily implies that we have on $C(m,k)$:

$$
\begin{aligned}
&\frac{1}{(2\pi)^2}\Bigg\{\left(\int |\mathcal{F}f(-u)|\left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right||\varphi(u)|\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|\mathrm{d}u\right)^2 \\
&\wedge \Bigg(\int |\varphi_Y(x)|^2\,\mathrm{d}x \\
&\quad \int |\mathcal{F}f(-u)|^2\left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2|\varphi(u)|^2\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|^2\mathrm{d}u\Bigg)\Bigg\} \\
&\leq \left(\left(\frac{5}{2}\kappa\right)(\log n)^{1/2} + \lambda_{f_{m,k}}\right)^2 \\
&\quad \frac{n^{-1}}{(2\pi)^2}\Bigg\{C_1\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|^2\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|^2 w(u)^{-2}\,\mathrm{d}u \\
&\qquad \wedge C_2\left(\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right| w(u)^{-1}\,\mathrm{d}u\right)^2\Bigg\} \\
&\leq \left(\left(\frac{5}{2}\kappa\right)(\log n)^{1/2} + \lambda_{f_{m,k}}\right)^2 n^{-1}\widetilde{\sigma}^2_{m,k}.
\end{aligned}
\tag{3.3.51}
$$

We use the trivial observation that we always have $x^2_{f_{m,k}} \leq \widetilde{\sigma}^2_{m,k}$ and hence $\lambda^2_{f_{m,k}} \leq c_1\widetilde{\lambda}^2_{m,k}$ as well as the fact that, by definition,

$$
\widetilde{\mathrm{H}}^2(m,k) \geq 8\left(\frac{5}{2}\kappa(\log n)^{1/2} + \sqrt{c_1}\widetilde{\lambda}_{m,k}\right)^2 n^{-1}\widetilde{\sigma}^2_{m,k}, \tag{3.3.52}
$$

to conclude that the last line is smaller than $\frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)$. We have thus shown that

$$\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}}\left\{\left|\left(\widetilde{\theta}_k-\widetilde{\theta}_m\right)-(\theta_k-\theta_m)\right|^2-\frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ 1(C(m,k))\right]=0. \tag{3.3.53}$$

It remains to show that the remainder term is negligible. The definition of $\frac{1}{\check{\varphi}_n}$ implies that we always have $\frac{1}{|\check{\varphi}_n|^2}\leq\kappa^{-2}(\log n)^{-1}n$. We can thus estimate

$$\begin{aligned}
&\left|\left(\widetilde{\theta}_k-\widetilde{\theta}_m\right)-(\theta_k-\theta_m)\right|^2\\
\leq&\frac{1}{2\pi^2}\left\{C_1\int|\mathcal{F}f(-u)|^2\left|\frac{1}{\check{\varphi}_n(u)}-\frac{1}{\varphi(u)}\right|^2|\varphi(u)|^2\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|^2\mathrm{d}u\right.\\
&\left.\wedge C_2\left(\int|\mathcal{F}f(-u)|\left|\frac{1}{\check{\varphi}_n(u)}-\frac{1}{\varphi(u)}\right||\varphi(u)|\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|\mathrm{d}u\right)^2\right\}\\
\leq&\frac{1}{2\pi^2}\left(\kappa^{-1}\left(n/\log n\right)^{\frac{1}{2}}+1\right)^2\left\{C_1\int|\mathcal{F}f(-u)|^2\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|^2\mathrm{d}u\right.\\
&\qquad\left.\wedge C_2\left(\int|\mathcal{F}f(-u)|\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|\mathrm{d}u\right)^2\right\}\\
=&\frac{1}{2\pi^2}\left(\kappa^{-1}(\log n)^{-1/2}n^{1/2}+1\right)^2x_{f_{m,k}}^2,
\end{aligned} \tag{3.3.54}$$

with $x_{f_{m,k}}$ defined as in Lemma 3.3.8. This implies

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\substack{k>m\\k\in\mathbb{N}}}\left\{\left|\left(\widetilde{\theta}_k-\widetilde{\theta}_m\right)-(\theta_k-\theta_m)\right|^2-\frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ 1\left(C(m,k)^c\right)\right]\\
\leq&\sum_{\substack{k>m\\k\in\mathbb{N}}}\mathbb{E}\left[\left\{\left|\left(\widetilde{\theta}_k-\widetilde{\theta}_m\right)-(\theta_k-\theta_m)\right|^2-\frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ 1\left(C(m,k)^c\right)\right]\\
\leq&\sum_{\substack{k>m\\k\in\mathbb{N}}}\frac{1}{2\pi^2}\left(\kappa^{-1}(\log n)^{-1/2}n^{1/2}+1\right)^2x_{f_{m,k}}^2\,\mathbb{P}\left(C(m,k)^c\right).
\end{aligned} \tag{3.3.55}$$

Now, Lemma 3.3.9 and the choice of κ imply that

$$\mathbb{P}\left(C(m,k)^c\right)=Cn^{-2}x_{f_{m,k}}^{-2}(m-k)^{-2}, \tag{3.3.56}$$

so the sum appearing in the last line is readily negligible.

This completes the proof of part (i).

(ii) We set, this time

$$\widetilde{\theta}_m:=\frac{1}{2\pi}\int\mathcal{F}f(-u)\frac{\varphi'(u)}{\check{\varphi}_n(u)}\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\mathrm{d}u \tag{3.3.57}$$

and use the same estimate as in formula (3.3.39). The considerations for the expression appearing in the last line of formula (3.3.39) are the same, line for line, as in the proof of part (i). We only have to replace φ_Z by φ' and use the fact that we have $\frac{\varphi'}{\varphi} = \Psi'$ instead of $\frac{\varphi_Z}{\varphi} = \varphi_Y$.

The only difficulty which arises in the Lévy model as compared to the deconvolution model is the fact that the quantity $\widehat{\varphi}_n'$ which appears in the definition of the estimator $\widehat{\theta}_m$ is unbounded. To be able to apply the Bernstein inequality, we have to introduce some truncation and then see that the remainder term is negligible.

Let

$$\bar{Z}_j := Z_j 1\left(\left\{|Z_j| \leq \frac{4}{\eta}\left(\log n + \log \widetilde{x}_{m,k}(k-m)\right)\right\}\right) \tag{3.3.58}$$

and $Z_j^r := Z_j - \bar{Z}_j$. Then we can decompose

$$\begin{aligned}
&\left|(\widehat{\theta}_k - \widehat{\theta}_m) - (\widetilde{\theta}_k - \widetilde{\theta}_m)\right|^2 \\
&= \left|\frac{1}{n}\sum_{j=1}^n \frac{1}{2\pi}\int \mathcal{F}f(-u)\frac{(Z_j e^{iuZ_j} - \mathbb{E}[Z_1 e^{iuZ_1}])}{\check{\varphi}_n(u)}\left(\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right)\mathrm{d}u\right|^2 \\
&\leq 2\left|\frac{1}{2\pi}\frac{1}{n}\sum_{j=1}^n \int \mathcal{F}f(-u)\frac{\bar{Z}_j e^{iuZ_j} - \mathbb{E}[\bar{Z}_1 e^{iuZ_1}]}{\check{\varphi}_n(u)}\left(\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right)\mathrm{d}u\right|^2 \\
&+ 2\left|\frac{1}{2\pi}\frac{1}{n}\sum_{j=1}^n \int \mathcal{F}f(-u)\frac{Z_j^r e^{iuZ_j^r} - \mathbb{E}[Z_1^r e^{iuZ_1^r}]}{\check{\varphi}_n(u)}\left(\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right)\mathrm{d}u\right|^2 \\
&=: 2\left|\left(\bar{\widehat{\theta}}_k - \bar{\widehat{\theta}}_m\right) - \left(\bar{\widetilde{\theta}}_k - \bar{\widetilde{\theta}}_m\right)\right|^2 + 2\left|\left(\widehat{\theta}_k^r - \widehat{\theta}_m^r\right) - \left(\widetilde{\theta}_k^r - \widetilde{\theta}_m^r\right)\right|^2.
\end{aligned} \tag{3.3.59}$$

Again, we find that

$$\mathbb{E}\left[\left|\left(\bar{\widehat{\theta}}_k - \bar{\widehat{\theta}}_m\right) - \left(\bar{\widetilde{\theta}}_k - \bar{\widetilde{\theta}}_m\right)\right|^2 \middle| \widehat{\varphi}_n\right] \leq \frac{1}{n}\widetilde{\sigma}_{m,k}^2 \ a.s. \tag{3.3.60}$$

and we can estimate, this time

$$\begin{aligned}
&\left\|\frac{1}{2\pi}\int \mathcal{F}f(-u)\frac{\bar{Z}_1 e^{iuZ_1}}{\check{\varphi}_n(u)}\left(\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right)\mathrm{d}u\right\|_\infty \\
\leq\ & \frac{4}{\eta}(\log(n\widetilde{x}_{m,k}(k-m)))\frac{1}{2\pi}\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right) - \mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|\mathrm{d}u \\
\leq\ & \sqrt{n}\,\frac{4}{\eta}(\log(n\widetilde{x}_{m,k}(k-m)))\widetilde{x}_{m,k} \quad a.s.
\end{aligned} \tag{3.3.61}$$

Another application of Lemma 3.3.2 gives

$$\mathbb{E}\left[\left\{\left|\left(\bar{\widehat{\theta}}_k - \bar{\widehat{\theta}}_m\right) - \left(\bar{\widetilde{\theta}}_k - \bar{\widetilde{\theta}}_m\right)\right|^2 - \frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+ \middle| \widehat{\varphi}_n\right]$$

$$\begin{aligned}
&\leq \quad 32\frac{\widetilde{\sigma}^2_{m,k}}{n}\exp\left(-\frac{c^{\text{pen}}\widetilde{\lambda}^2_{m,k}\widetilde{\sigma}^2_{m,k}}{64\widetilde{\sigma}^2_{m,k}}\right)\\
&+ \quad 128\sqrt{2}\frac{16}{\eta^2}\frac{(\log(n\widetilde{x}_{m,k}(k-m)))^2\widetilde{x}^2_{m,k}}{n}\\
&\qquad\qquad \exp\left(-\frac{c^{\text{pen}}\widetilde{\lambda}_{m,k}\widetilde{x}_{m,k}}{64\frac{4}{\eta}\left(\log(n\widetilde{x}_{m,k}(k-m))\right)\widetilde{x}_{m,k}}\right)\ a.s. \qquad (3.3.62)
\end{aligned}$$

Using, once again, the fact that $c^{\text{pen}} \geq 64$ and $\widetilde{\lambda}^2_{m,k} \geq \log\left(\widetilde{\sigma}^2_{m,k}(k-m)^2\right)$ and using the fact that, this time,

$$\widetilde{\lambda}_{m,k} \geq \frac{8}{\eta}\left(\log(n\widetilde{x}_{m,k}(k-m))\right)\log\left(\log(n\widetilde{x}_{m,k}(k-m))\right)^2\log\left(\widetilde{x}^2_{m,k}(k-m)^2\right)$$

(see line 3.2.56), we can continue by estimating

$$\begin{aligned}
&32\frac{\widetilde{\sigma}^2_{m,k}}{n}\exp\left(-\frac{c^{\text{pen}}\widetilde{\lambda}^2_{m,k}\widetilde{\sigma}^2_{m,k}}{64\widetilde{\sigma}^2_{m,k}}\right)\\
&+2048\frac{\sqrt{2}}{\eta^2}\frac{(\log(n\widetilde{x}_{m,k}(k-m)))^2\widetilde{x}^2_{m,k}}{n}\exp\left(-\frac{c^{\text{pen}}\widetilde{\lambda}_{m,k}\widetilde{x}_{m,k}}{\frac{256}{\eta}\left(\log(n\widetilde{x}_{m,k}(k-m))\right)\widetilde{x}_{m,k}}\right)\\
&\leq 32\frac{\widetilde{\sigma}^2_{m,k}}{n}\exp\left(-\widetilde{\lambda}^2_{m,k}\right)\\
&+128\sqrt{2}\frac{16}{\eta^2}\frac{(\log(n\widetilde{x}_{m,k}(k-m)))^2\widetilde{x}^2_{m,k}}{n}\exp\left(-\frac{\widetilde{\lambda}_{m,k}}{\frac{4}{\eta}\log(n\widetilde{x}_{m,k}(k-m))}\right)\\
&\leq 32\frac{\widetilde{\sigma}^2_{m,k}}{n}\widetilde{\sigma}^{-2}_{m,k}(k-m)^{-2}\\
&+128\sqrt{2}\frac{16}{\eta^2}\frac{(\log(n\widetilde{x}_{m,k}(k-m)))^2\widetilde{x}^2_{m,k}}{n}(\log(n\widetilde{x}_{m,k}(k-m)))^{-2}\widetilde{x}^{-2}_{m,k}(k-m)^{-2}.
\end{aligned}$$

We have thus shown that, almost surely,

$$\begin{aligned}
&\sum_{\substack{k\geq m\\ k\in\mathbb{N}}}\mathbb{E}\left[\left\{\left|\left(\widehat{\overline{\theta}}_k-\widehat{\overline{\theta}}_m\right)-\left(\widetilde{\overline{\theta}}_k-\widetilde{\overline{\theta}}_m\right)\right|^2-\frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\middle|\widehat{\varphi}_n\right]\\
&\leq \quad (32+128\sqrt{2}\frac{16}{\eta^2})n^{-1}\sum_{\substack{k\geq m\\ k\in\mathbb{N}}}(k-m)^{-2}. \qquad (3.3.63)
\end{aligned}$$

It is important to recall, at this point, that we have split our observations into two independent samples, which implies that $\widehat{\varphi}'_n$ and $\widehat{\varphi}_n$ are independent. This allows to apply the Bernstein inequality.

The remainder term can be estimated as follows:

$$\mathbb{E}\left[\left|\left(\widehat{\theta}^r_k-\widehat{\theta}^r_m\right)-\left(\widetilde{\theta}^r_k-\widetilde{\theta}^r_m\right)\right|^2\middle|\widehat{\varphi}_n\right]$$

$$\leq \frac{1}{n}\,\mathbb{E}\left[\left|\frac{1}{2\pi}\int \mathcal{F}f(-u)\frac{Z_1^r e^{iuZ_1}}{\check{\varphi}_n(u)}\left(\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right)\mathrm{d}u\right|^2\middle|\widehat{\varphi}_n\right]$$
$$\leq \frac{1}{n}\frac{1}{(2\pi)^2}\,\mathbb{E}\left[|Z_1^r|^2|\widehat{\varphi}_n\right]\left(\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|\left|\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right|\mathrm{d}u\right)^2$$
$$= \widetilde{x}_{m,k}^2\,\mathbb{E}\left[|Z_1^r|^2|\widehat{\varphi}_n\right]\ a.s. \tag{3.3.64}$$

Now, we apply Markov's inequality to find that

$$\begin{aligned}
&\mathbb{E}\left[|Z_1^r|^2|\widehat{\varphi}_n\right]\\
=\ &\mathbb{E}\left[|Z_1|^2\,1\left(\left\{|Z_1|>\frac{4}{\eta}\log\left(nx_{m,k}(k-m)\right)\right\}\right)|\widehat{\varphi}_n\right]\\
\leq\ &\mathbb{E}\left[|Z_1|^2\exp\left(\frac{\eta}{2}|Z_1|\right)\middle|\widehat{\varphi}_n\right]\exp\left(-2\log\left(n\widetilde{x}_{m,k}(k-m)\right)\right)\\
\leq\ &n^{-1}\frac{4}{\eta^2}\,\mathbb{E}\left[\exp\left(\eta|Z_1|\right)\right]\widetilde{x}_{m,k}^{-2}(k-m)^{-2}a.s.
\end{aligned}\tag{3.3.65}$$

We have thus shown that

$$\begin{aligned}
&\sum_{\substack{k\geq m\\k\in\mathbb{N}}}\mathbb{E}\left[\left|\left(\widehat{\theta}_k^r-\widehat{\theta}_m^r\right)-\left(\widetilde{\theta}_k^r-\widetilde{\theta}_m^r\right)\right|^2\middle|\widehat{\varphi}_n\right]\\
\leq\ &\frac{4}{\eta^2}\,\mathbb{E}\left[\exp\left(\eta|Z_1|\right)\right]n^{-1}\sum_{\substack{k\geq m\\k\in\mathbb{N}}}(k-m)^{-2}a.s.
\end{aligned}\tag{3.3.66}$$

Together with (3.3.63), this gives for some universal constant C':

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\substack{k\geq m\\k\in\mathbb{N}}}\left\{\left|\left(\widehat{\theta}_k-\widehat{\theta}_m\right)-\left(\widetilde{\theta}_k-\widetilde{\theta}_m\right)\right|^2-\frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\right]\\
\leq\ &\mathbb{E}\left[\sum_{\substack{k\geq m\\k\in\mathbb{N}}}\mathbb{E}\left[\left\{\left|\left(\widehat{\theta}_k-\widehat{\theta}_m\right)-\left(\widetilde{\theta}_k-\widetilde{\theta}_m\right)\right|^2-\frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\middle|\widehat{\varphi}_n\right]\right]\\
=\ &C'n^{-1},
\end{aligned}\tag{3.3.67}$$

and hence the statement of part (ii).

(iii) This part of the statement differs from part (ii) in the fact that the truncation introduced in (3.3.58) is now different. We set, this time

$$\bar{Z}_j := Z_j 1\left(\{|Z_j|\leq\sqrt{n}\}\right). \tag{3.3.68}$$

Then we can estimate

$$\left\|\frac{1}{2\pi}\int\mathcal{F}f(-u)\frac{\bar{Z}_1e^{iuZ_1}}{\check{\varphi}(u)}\left(\mathcal{F}\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\mathrm{K}\left(\frac{u}{m}\right)\right)\mathrm{d}u\right\|_\infty$$

$$
\begin{aligned}
&\le \sqrt{n}\frac{1}{2\pi}\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|\left|\mathcal{F}\,\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\,\mathrm{K}\left(\frac{u}{m}\right)\right|\,\mathrm{d}u\\
&\le \sqrt{n}\,\widetilde{\sigma}_{m,k}. && (3.3.69)
\end{aligned}
$$

The last inequality is a consequence of the fact that $\bar{C}_1=\infty$ and $\bar{C}_2\ge 1$. Again, we apply the integral version of the Bernstein inequality to the truncated random variables and find that, almost surely,

$$
\begin{aligned}
&\mathbb{E}\left[\left\{\left|\left(\widehat{\widetilde{\theta}}_k-\widehat{\widetilde{\theta}}_m\right)-\left(\widetilde{\widetilde{\theta}}_k-\widetilde{\widetilde{\theta}}_m\right)\right|^2-\frac{1}{8}\widetilde{\mathrm{H}}^2(m,k)\right\}_+\middle|\widehat{\varphi}_n\right]\\
&= O\left(n^{-1}(k-m)^{-2}\right). && (3.3.70)
\end{aligned}
$$

The remainder term can now be estimated as follows: As in the proof of part (ii), we have

$$
\begin{aligned}
&\mathbb{E}\left[\left|\left(\widehat{\theta}^r_k-\widehat{\theta}^r_m\right)-\left(\widetilde{\theta}^r_k-\widetilde{\theta}^r_m\right)\right|^2\middle|\widehat{\varphi}_n\right]\\
&\le\frac{1}{n}\,\mathbb{E}\left[\left|\frac{1}{2\pi}\int\mathcal{F}f(-u)\frac{Z_1^r e^{iuZ_1}}{\check{\varphi}_n(u)}\left(\mathcal{F}\,\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\,\mathrm{K}\left(\frac{u}{m}\right)\right)\,\mathrm{d}u\right|^2\middle|\widehat{\varphi}_n\right]\\
&\le\frac{1}{n}\frac{1}{(2\pi)^2}\,\mathbb{E}\left[|Z_1^r|^2|\widehat{\varphi}_n\right]\left(\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|\left|\mathcal{F}\,\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\,\mathrm{K}\left(\frac{u}{m}\right)\right|\,\mathrm{d}u\right)^2 && (3.3.71)
\end{aligned}
$$

We have now, by assumption, $m_n\le n$ and thus $k,m\le n$. Moreover, we have $\frac{1}{\check{\varphi}_n(u)}\le n^{1/2}$ by construction, $\|f\|_\infty<\infty$ by assumption and the Fourier transforms of the kernels are supported in $[-\pi n,\pi n]$. We can thus continue by estimating

$$
\begin{aligned}
&\frac{1}{n}\frac{1}{(2\pi)^2}\,\mathbb{E}\left[|Z_1^r|^2|\widehat{\varphi}_n\right]\left(\int\left|\frac{\mathcal{F}f(-u)}{\check{\varphi}_n(u)}\right|\left|\mathcal{F}\,\mathrm{K}\left(\frac{u}{k}\right)-\mathcal{F}\,\mathrm{K}\left(\frac{u}{m}\right)\right|\,\mathrm{d}u\right)^2\\
\le\;&\frac{1}{n}\frac{1}{(2\pi)^2}\,\mathbb{E}\left[|Z_1^r|^2|\widehat{\varphi}_n\right]2\pi n^3\|f\|_\infty. && (3.3.72)
\end{aligned}
$$

Next, Markov's inequality gives

$$
\mathbb{E}\left[|Z_1^r|^2|\widehat{\varphi}_n\right]=\mathbb{E}\left[|Z_1|^2 1\left(\{|Z_1|>\sqrt{n}\}\right)\right]\le n^{-4}\,\mathbb{E}\left[|Z_1|^{10}\right]. \qquad (3.3.73)
$$

We have thus shown that

$$
\begin{aligned}
&\sum_{\substack{k\ge m\\ k\in\mathcal{M}}}\mathbb{E}\left[\left|\left(\widehat{\theta}^r_k-\widehat{\theta}^r_m\right)-\left(\theta^r_k-\theta^r_m\right)\right|^2\right]\\
\le\;&\sum_{\substack{k\ge m\\ k\le n}}n^{-2}\frac{1}{2\pi}\,\mathbb{E}\left[|Z_1|^{10}\right]\le\frac{1}{(2\pi)^2}\,\mathbb{E}\left[|Z_1|^{10}\right]n^{-1}. && (3.3.74)
\end{aligned}
$$

and hence the statement for the remainder term, which completes the proof of part (iii). $\square$

We are now ready to prove Theorem 3.2.10 and Theorem 3.2.11 and Theorem 3.2.12. The proofs of the three statements are completely analogous.

Proof of Theorem 3.2.10, Theorem 3.2.11 and Theorem 3.2.12

In what follows, let m^* denote the oracle cutoff, that is,

$$\begin{aligned} m^* &= \operatorname*{argmin}_{m\in\mathcal{M}} \mathrm{Crit}(m) \\ &:= \operatorname*{argmin}_{m\in\mathcal{M}} \left\{ \sup_{\substack{k\geq m\\ k\in\mathcal{M}}} |\theta_k - \theta_m|^2 + \mathrm{pen}(m) \right\}. \end{aligned} \tag{3.3.75}$$

We start by considering the loss on the set $\{\widehat{m} \leq m^*\}$.

We use the estimate

$$\begin{aligned} &\left|\theta - \widehat{\theta}_{\widehat{m}}\right|^2 1\left(\{\widehat{m} \leq m^*\}\right) \\ \leq\ & 2\left|\theta - \widehat{\theta}_{m^*}\right|^2 1\left(\{\widehat{m} \leq m^*\}\right) + 2\left|\widehat{\theta}_{m^*} - \widehat{\theta}_{\widehat{m}}\right| 1\left(\{\widehat{m} \leq m^*\}\right). \end{aligned} \tag{3.3.76}$$

First of all, Lemma 3.2.6 and the definition of the penalty term imply that

$$\begin{aligned} \mathbb{E}\left[|\theta - \widehat{\theta}_{m^*}|^2\right] &\leq 2|\theta - \theta_{m^*}|^2 + 2\,\mathbb{E}\left[\left|\theta_{m^*} - \widehat{\theta}_{m^*}\right|^2\right] \\ &\leq 2|\theta - \theta_{m^*}|^2 + 2C_{\aleph}\,\mathrm{pen}(m^*). \end{aligned} \tag{3.3.77}$$

Next, the definition of $\widehat{m}$ gives

$$\begin{aligned} &\left|\widehat{\theta}_{m^*} - \widehat{\theta}_{\widehat{m}}\right|^2 1\left(\{\widehat{m} \leq m^*\}\right) \\ &\leq \widetilde{\mathrm{pen}}(m^*) + \sup_{\substack{k>m^*\\ k\in\mathcal{M}}} \left\{ \left|\widehat{\theta}_k - \widehat{\theta}_{m^*}\right|^2 - \widetilde{\mathrm{H}}^2(m^*,k) \right\} + \widetilde{\mathrm{H}}^2(\widehat{m},m^*) 1\left(\{\widehat{m} \leq m^*\}\right). \end{aligned} \tag{3.3.78}$$

We can estimate

$$\begin{aligned} &\sup_{\substack{k>m^*\\ k\in\mathcal{M}}} \left\{ \left|\widehat{\theta}_k - \widehat{\theta}_{m^*}\right|^2 - \widetilde{\mathrm{H}}^2(m^*,k) \right\} \\ &\leq \sup_{\substack{k>m^*\\ k\in\mathcal{M}}} \left\{ 2\left|(\widehat{\theta}_k - \widehat{\theta}_{m^*}) - (\theta_k - \theta_{m^*})\right|^2 - \widetilde{\mathrm{H}}^2(m^*,k) \right\} + 2 \sup_{\substack{k>m^*\\ k\in\mathcal{M}}} |\theta_k - \theta_{m^*}|^2. \end{aligned} \tag{3.3.79}$$

Proposition 3.3.10 readily implies that for some positive constant C, we have

$$\mathbb{E}\left[\sup_{\substack{k>m^*\\ k\in\mathcal{M}}} \left\{ 2\left|(\widehat{\theta}_k - \widehat{\theta}_{m^*}) - (\theta_k - \theta_{m^*})\right|^2 - \widetilde{\mathrm{H}}^2(m^*,k) \right\} \right] \leq Cn^{-1}. \tag{3.3.80}$$

Next, we observe that, by definition of $\widetilde{\mathrm{H}}^2$ and $\widetilde{\mathrm{pen}}$,

$$\widetilde{\mathrm{H}}(\widehat{m}, m^*) 1\left(\{\widehat{m} \leq m^*\}\right) \leq \widetilde{\mathrm{pen}}(m^*). \tag{3.3.81}$$

Since we have chosen $\kappa \geq 2\left(\sqrt{4c_1} + \gamma\right)$, we can apply Corollary 3.3.7 to find that for some positive constant C_{κ} depending only on the choice of the constants,

$$\mathbb{E}\left[\widetilde{\mathrm{pen}}(m^*)\right] \leq C_{\kappa}\, \mathrm{pen}(m^*). \tag{3.3.82}$$

To do this, we apply the Cauchy-Schwarz inequality to see that

$$\mathbb{E}\left[\widetilde{\lambda}^2_{m^*}\widetilde{\sigma}^2\right] \leq \left(\mathbb{E}\left[\widetilde{\lambda}^4_{m^*}\right]\right)^{1/2}\left(\mathbb{E}\left[\widetilde{\sigma}^4_{m^*}\right]\right)^{1/2} \tag{3.3.83}$$

and then use Corollary 3.3.7 to derive that

$$\left(\mathbb{E}\left[\widetilde{\lambda}^4_{m^*}\right]\right)^{1/2}\left(\mathbb{E}\left[\widetilde{\sigma}^4_{m^*}\right]\right)^{1/2} \leq C_{\kappa}\lambda^2_{m^*}\sigma^2_{m^*}. \tag{3.3.84}$$

Putting the above results together, we have shown that for some positive constant C_{κ} depending only on the choice of the constants and some constant C specified in Proposition 3.3.10,

$$\begin{aligned}
&\mathbb{E}\left[\left|\theta - \widehat{\theta}_{\widehat{m}}\right|^2 1\left(\{\widehat{m} \leq m^*\}\right)\right] \\
&\leq 2\left|\theta - \theta_{m^*}\right|^2 + C_{\kappa}\, \mathrm{pen}(m^*) + 2 \sup_{\substack{k > m^* \\ k \in \mathcal{M}}} \left|\theta_k - \theta_{m^*}\right|^2 + Cn^{-1} \\
&\leq (6 + C_{\kappa}) \inf_{m \in \mathcal{M}} \left\{\left|\theta - \theta_{m_n}\right|^2 + \sup_{\substack{k \geq m \\ k \in \mathcal{M}}} \left|\theta_k - \theta_m\right|^2 + \mathrm{pen}(m)\right\} + C' n^{-1}.
\end{aligned} \tag{3.3.85}$$

This is the desired result for the expected loss on $\{\widehat{m} \leq m^*\}$.

It remains to consider the loss on the set $\{\widehat{m} > m^*\}$.

We use the estimate

$$|\theta - \widehat{\theta}_{\widehat{m}}|^2 \leq 3|\theta - \theta_{m^*}|^2 + 3|\theta_{\widehat{m}} - \theta_{m*}|^2 + 3|\theta_{\widehat{m}} - \widehat{\theta}_{\widehat{m}}|^2. \tag{3.3.86}$$

First, we clearly have

$$\begin{aligned}
&3\left(|\theta - \theta_{m^*}|^2 + |\theta_{\widehat{m}} - \theta_{m^*}|^2\right) 1\left(\{\widehat{m} > m^*\}\right) \\
\leq\ & 3|\theta - \theta_{m^*}|^2 + 3 \sup_{\substack{k > m* \\ k \in \mathcal{M}}} |\theta_k - \theta_{m*}|^2.
\end{aligned} \tag{3.3.87}$$

Next, we can estimate

$$|\theta_{\widehat{m}} - \widehat{\theta}_{\widehat{m}}|^2 1\left(\{\widehat{m} > m^*\}\right) = \sum_{\substack{k > m^* \\ k \in \mathcal{M}}} |\theta_k - \widehat{\theta}_k|^2 1\left(\{\widehat{m} = k\}\right)$$

$$\leq \sum_{\substack{k<m^* \\ k\in\mathcal{M}}} \left\{|\theta_k - \widehat{\theta}_k|^2 - \widetilde{\text{pen}}(k)\right\}_+ + \sum_{\substack{k>m^* \\ k\in\mathcal{M}}} \widetilde{\text{pen}}(k)1\left(\{\widehat{m}=k\}\right). \quad (3.3.88)$$

Taking expectation and applying again Proposition 3.3.10, we find that

$$\sum_{\substack{k>m^* \\ k\in\mathcal{M}}} \mathbb{E}\left[\left\{|\theta_k - \widehat{\theta}_k|^2 - \widetilde{\text{pen}}(k)\right\}_+\right] \leq Cn^{-1}. \quad (3.3.89)$$

Moreover, by definition of $\widehat{m}$, we have on $\{\widehat{m}=k\}$:

$$\begin{aligned}
&\widetilde{\text{pen}}(k) \\
&\leq \widetilde{\text{pen}}(m^*) + \sup_{\substack{l>m^* \\ l\in\mathcal{M}}} \left\{|\widehat{\theta}_l - \widehat{\theta}_{m^*}|^2 - \widetilde{\mathrm{H}}^2(m^*,l)\right\}_+ \\
&\leq \widetilde{\text{pen}}(m^*) \\
&+2\sup_{\substack{l>m^* \\ l\in\mathcal{M}}} \left\{\left|(\widehat{\theta}_l - \widehat{\theta}^*_m) - (\theta_l - \theta_{m^*})\right|^2 - \frac{1}{2}\widetilde{\mathrm{H}}^2(m^*,l)\right\}_+ + 2\sup_{\substack{l>m^* \\ l\in\mathcal{M}}} |\theta_l - \theta_{m^*}|^2. \quad (3.3.90)
\end{aligned}$$

Again, one can deal with the first expression in the last line of formula (3.3.90), using Proposition 3.3.10. Moreover, we can argue, once again, that by Corollary 3.3.7, we have for some positive constant $C_{\varkappa}$:

$$\mathbb{E}\left[\widetilde{\text{pen}}(m^*)\right] \leq C_{\varkappa}\,\text{pen}(m^*). \quad (3.3.91)$$

We have thus shown that

$$\begin{aligned}
&\mathbb{E}\left[|\theta_{\widehat{m}} - \widehat{\theta}_{\widehat{m}}|^2 1\left(\{\widehat{m}>m^*\}\right)\right] \\
\leq\ & 2\sup_{\substack{l>m^* \\ l\in\mathcal{M}}} |\theta_l - \theta_{m^*}|^2 + C_{\varkappa}\,\text{pen}(m^*) + 2Cn^{-1}. \quad (3.3.92)
\end{aligned}$$

Together with (3.3.86) and (3.3.87), this gives

$$\begin{aligned}
&\mathbb{E}\left[|\theta - \widehat{\theta}_{\widehat{m}}|^2 1\left(\{\widehat{m}>m^*\}\right)\right] \\
&\leq (11+C_{\varkappa}) \inf_{m\in\mathcal{M}} \left\{ |\theta - \theta_{m_n}|^2 + \sup_{\substack{k>m^* \\ k\in\mathcal{M}}} |\theta_k - \theta_{m^*}|^2 + \text{pen}(m) \right\} + 2Cn^{-1}, \quad (3.3.93)
\end{aligned}$$

which is the desired result for $\{\widehat{m}>m^*\}$. This completes the proof. □

3.3.2 Proof of the main result of Section 3.2.1

In what follows, we formulate some auxiliary results for the global threshold estimator which has been introduced in Section 3.2.1.

Indeed, these results can be treated as special cases of the more general results which have been obtained in the preceding section. For this reason, we only sketch the proofs to avoid redundancy.

Lemma 3.3.11. *For $\tau > 0$, let*

$$U_n := U_n^\tau := \inf\left\{u \in \mathbb{R}^+ : |\varphi(u)| \le 3\tau(\log n)^{1/2} w(u)^{-1} n^{-1/2}\right\} \tag{3.3.94}$$

and

$$U_n' := U_n^{\tau\prime} := \inf\left\{u \in \mathbb{R}^+ := |\varphi(u)| \le \tau(\log n)^{1/2} w(u)^{-1} n^{-1/2}\right\}. \tag{3.3.95}$$

Moreover, let

$$\widehat{U}_n := \widehat{U}_n^\tau := \inf\left\{u \in \mathbb{R}^+ : |\widehat{\varphi}_n(u)| \le 2\tau(\log n)^{\frac{1}{2}} w(u)^{-1} n^{-1/2}\right\}. \tag{3.3.96}$$

Then we have for some positive constant C_* depending on the choice of γ and τ:

$$\mathbb{P}\left(\left\{\widehat{U}_n < U_n\right\} \cup \left\{\widehat{U}_n > U_n'\right\}\right) \le C_{**} n^{-\frac{(\tau-\gamma)^2}{c_1}}. \tag{3.3.97}$$

Proof. This is an immediate consequence of Lemma 3.3.4. □

Next, we can prove the following versions of Proposition 3.3.5 and Corollary 3.3.6:

Lemma 3.3.12. *Let $\frac{1}{\check{\varphi}_n} = \frac{1}{\check{\varphi}_n^{\kappa,\delta}}$ be defined according to Definition 3.2.1. Assume that for some $\gamma > 0$ and some $p > 0$, we have $\kappa \ge 2(\sqrt{pc_1} + \gamma)$, where c_1 denotes the constant in Talagrands inequality. Then we find that for some constant C_* depending only on the choice of γ and δ,*

$$\mathbb{P}\left(\left\{\exists u \in \mathbb{R} : \left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2 > \frac{C_\varphi'^2}{C_\varphi^2}\left((4\kappa)^2 \frac{w(u)^{-1}\frac{\log n}{n}}{|\varphi(u)|^4} \wedge \left(\frac{5}{2}\right)^2 \frac{1}{|\varphi(u)|^2}\right)\right\}\right)$$
$$\le C_* n^{-p}. \tag{3.3.98}$$

Proof. It is enough to consider the set

$$C := \left\{\widehat{U}_n \in [U_n, U_n']\right\}. \tag{3.3.99}$$

We introduce, this time, the partition $\mathbb{R} = \mathbb{R}_1 \cup \mathbb{R}_2 \cup \mathbb{R}_3$ with

$$\mathbb{R}_1 := \{u \in [-U_n, U_n]\}, \quad \mathbb{R}_2 = \{u \in [-U_n', U_n']^c\} \tag{3.3.100}$$

and

$$\mathbb{R}_3 := \{u \in [-U_n', -U_n] \cup [U_n, U_n']\}. \tag{3.3.101}$$

We use the fact that on the set C, we have on $\mathbb{R}_1$,

$$|\varphi(u)| \ge \frac{3}{2}\kappa(\log n)^{1/2} w(u)^{-1} n^{-1/2}, \tag{3.3.102}$$

as well as

$$\left|\frac{1}{\check{\varphi}_n(u)}\right| = \left|\frac{1}{\widehat{\varphi}_n(u)}\right| \le \kappa^{-1}(\log n)^{1/2} w(u) n^{1/2}, \tag{3.3.103}$$

on $\mathbb{R}_2$,

$$|\varphi(u)| \leq \frac{C'_\varphi}{C_\varphi}\frac{\kappa}{2}(\log n)^{1/2} w(u)^{-1} n^{-1/2} \tag{3.3.104}$$

and

$$\frac{1}{\check{\varphi}_n(u)} = 0 \tag{3.3.105}$$

and finally, on $\mathbb{R}_3$,

$$\frac{\kappa}{2}(\log n)^{1/2} w(u)^{-1} n^{-1/2} \leq |\varphi(u)| \leq \frac{C'_\varphi 3\kappa}{C_\varphi 2}(\log n)^{1/2} n^{-1/2}, \tag{3.3.106}$$

and

$$\left|\frac{1}{\check{\varphi}_n(u)}\right| \leq \kappa^{-1}(\log n)^{-1} n^{1/2}. \tag{3.3.107}$$

Combining these inequalities gives the desired result on the set C. The remainder term is negligible thanks to Lemma 3.3.11. □

Corollary 3.3.13. *In the situation of the preceding statement, we have*

$$\mathbb{P}\left(\left\{\exists u \in [-\widehat{U}_n, \widehat{U}_n] : \left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2 > (4\kappa)^2 \frac{(\log n) w(u)^{-1} n^{-1}}{|\check{\varphi}_n(u)|^2 |\varphi(u)|^2}\right\}\right)$$
$$\leq C_\kappa n^{-p}.$$

Proof. This result is derived from the preceding statement as Corollary 3.3.6 is derived from Proposition 3.3.5. □

In analogy with Lemma 3.2.6, we derive the following statement:

Lemma 3.3.14. *Let $\frac{1}{\check{\varphi}_n}$ be the estimator which has been introduced in Definition 3.2.1. Let κ be chosen such that for some $\gamma > 0$, we have $\kappa \geq 2(\sqrt{2c_1} + \gamma)$. Let Assumption 3.2.2 be satisfied. Then we have for some constant C_κ depending on the choice of κ, γ and δ:*

$$\mathbb{E}\left[\sup_{u \in \mathbb{R}} \frac{\left|\frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^2}{\frac{(\log n) w(u)^{-2} n^{-1}}{|\varphi(u)|^4} \wedge \frac{1}{|\varphi(u)|^2}}\right] \leq \frac{C'^2_\varphi}{C^2_\varphi} C_\kappa.$$

Proof. This result is derived directly from Lemma 3.3.12 by observing that the estimate trivially holds true on a favourable set C and the remainder term is negligible. □

Finally, we will need the following version of Proposition 3.3.10:

Proposition 3.3.15. *For arbitrary $m \in \mathbb{N}$, let $\widehat{g}_m = \widehat{g}_{m,n}$ be defined by (3.2.24). Let g_m be defined by (3.2.4). Let $\widetilde{\mathrm{H}}^2(m,k) := \widetilde{\mathrm{pen}}(k) - \widetilde{\mathrm{pen}}(m)$. For $k \in \mathcal{M}$, let k' denote the random index $k' = k \wedge \lfloor \frac{1}{\pi}\widehat{U}_n \rfloor$. Then we find that* *for any $m \in \mathcal{M}$:*

$$\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}}}\left\{\|(\widehat{g}_k-\widehat{g}_m)-(g_{k'}-g_{m'})\|^2_{\mathrm{L}^2}-\frac{1}{2}\widetilde{\mathrm{H}}^2(m,k)\right\}\right]\leq Cn^{-1}, \tag{3.3.108}$$

with some positive constant C depending only on the choice of κ and δ.

Proof. (sketch) We define

$$\widetilde{g}_m := \mathcal{F}^{-1}\left(\frac{\varphi_Z(u)}{\check{\varphi}(u)}1\left([-\pi m,\pi m]\right)(u)\right), \tag{3.3.109}$$

and consider, in analogy with the proof of Proposition 3.3.10,

$$\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}}}\left\{\|\left(\widehat{g}_k-\widehat{g}_m\right)-\left(\widetilde{g}_k-\widetilde{g}_m\right)\|^2_{\mathrm{L}^2}-\frac{1}{8}\widetilde{\mathrm{H}}(m,k)\right\}_+\right] \tag{3.3.110}$$

as well as

$$\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}}}\left\{\|\left(\widetilde{g}_k-\widetilde{g}_m\right)-\left(g_{k'}-g_{m'}\right)\|^2_{\mathrm{L}^2}-\frac{1}{8}\widetilde{\mathrm{H}}(m,k)\right\}_+\right]. \tag{3.3.111}$$

Again, we condition on the characteristic function and use the Bernstein inequality to deal with the quantity appearing in formula (3.3.110). Since the absolute value of the characteristic function $\widehat{\varphi}_Z$ is bounded above, we do not need here a finite moment assumption.

The quantity appearing in formula (3.3.111) is treated in the same way as the quantity appearing in the last line of formula (3.3.39) in the proof of Proposition 3.3.10. We introduce, this time, the favourable sets

$$\begin{aligned}&C(m,k)\\ &:=\left\{\forall u\in[-\widehat{U}_n,\widehat{U}_n]:\left|\frac{1}{\check{\varphi}_n(u)}-\frac{1}{\varphi(u)}\right|^2\leq\frac{\left(\frac{5}{2}\kappa(\log n)^{\frac{1}{2}}+\lambda_{g_{m,k}}\right)^2 w(u)^{-2}}{|\check{\varphi}_n(u)|^2\,|\varphi(u)|^2\,n}\right\},\end{aligned} \tag{3.3.112}$$

where we set, this time,

$$\lambda_{g_{m,k}} := \sqrt{c_1}\log\left(x^2_{g_{m,k}}(k-m)^2\right) \tag{3.3.113}$$

and

$$x^2_{g_{m,k}} := \frac{1}{2\pi}\int\limits_{\{\pi m\leq|u|\leq\pi k\}} 1\,\mathrm{d}u=(k-m). \tag{3.3.114}$$

We can estimate on $C(m,k)$:

$$
\begin{aligned}
&\|(\widetilde{g}_k - \widetilde{g}_m) - (g_{k'} - g_{m'})\|_{\mathrm{L}^2}^2 \\
&= \int\limits_{\{\pi m \le |u| \le \pi k\}} |\varphi_Z(u)|^2 \left| \frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right|^2 \mathrm{d}u \\
&= \int\limits_{\{\pi m \le |u| \le \pi k\}} |\varphi_Y(u)|^2 \, |\varphi(u)|^2 \left| \frac{1}{\check{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right|^2 \mathrm{d}u \\
&\le \left(\frac{5}{2}\kappa (\log n)^{\frac{1}{2}} + \lambda_{g_m,k} \right)^2 \int\limits_{\{\pi m \le |u| \le \pi k\}} \frac{w(u)^{-2}}{|\check{\varphi}_n(u)|^2} \mathrm{d}u \\
&\le \left(\frac{5}{2}\kappa (\log n)^{\frac{1}{2}} + \lambda_{m,k} \right)^2 \int\limits_{\{\pi m \le |u| \le \pi k\}} \frac{w(u)^{-2}}{|\check{\varphi}_n(u)|^2} \mathrm{d}u \\
&\le \frac{1}{8} \widetilde{\mathrm{H}}^2(m,k),
\end{aligned}
\tag{3.3.115}
$$

where the last estimate follows immediately from the definition of the penalty term.

Finally, we show that the remainder term is negligible, using the same arguments as in the proof of Proposition 3.3.10. □

We are now ready to prove the main result for the global threshold estimator:

Proof of Theorem 3.2.4

Let m^* denote the oracle cutoff. Again, we start by considering the set $\{\widehat{m} \le m^*\}$. We can estimate

$$
\|g - \widehat{g}_{\widehat{m}}\|_{\mathrm{L}^2}^2 \le 2\|g - \widehat{g}_{m^*}\|_{\mathrm{L}^2}^2 + 2\|\widehat{g}_{m^*} - \widehat{g}_{\widehat{m}}\|_{\mathrm{L}^2}^2. \tag{3.3.116}
$$

First of all, Lemma 3.3.14 allows to estimate

$$
\mathbb{E}\left[\|g - \widehat{g}_{m^*}\|_{\mathrm{L}^2}^2 \right] \le 2\|g - g_{m^*}\|_{\mathrm{L}^2}^2 + 2\frac{C_\varphi'^2}{C_\varphi^2} C_{\kappa} \,\mathrm{pen}(m^*). \tag{3.3.117}
$$

Next, we have

$$
\|\widehat{g}_{m^*} - \widehat{g}_{\widehat{m}}\|_{\mathrm{L}^2}^2 \mathbb{1}\left(\{\widehat{m} \le m^*\}\right) = \left(\|\widehat{g}_{m^*}\|_{\mathrm{L}^2}^2 - \|\widehat{g}_{\widehat{m}}\|_{\mathrm{L}^2}^2 \right) \mathbb{1}\left(\{\widehat{m} \le m^*\}\right). \tag{3.3.118}
$$

The definition of $\widehat{m}$ implies that

$$
-\|\widehat{g}_{\widehat{m}}\|_{\mathrm{L}^2}^2 \le -\|\widehat{g}_{m^*}\|_{\mathrm{L}^2}^2 + \widetilde{\mathrm{pen}}(m^*). \tag{3.3.119}
$$

From this we derive that

$$
\left(\|\widehat{g}_{m^*}\|_{\mathrm{L}^2}^2 - \|\widehat{g}_{\widehat{m}}\|_{\mathrm{L}^2}^2 \right) \mathbb{1}\left(\{\widehat{m} \le m^*\}\right) \le \widetilde{\mathrm{pen}}(m^*). \tag{3.3.120}
$$

We can extend Lemma 3.3.14 to see that for some positive constant $C_{\varkappa}$ depending only on the choice of the constants,

$$\mathbb{E}\left[\widetilde{\mathrm{pen}}(m^*)\right] \leq \frac{C'^2_{\varphi}}{C^2_{\varphi}} C_{\varkappa} \,\mathrm{pen}(m^*). \tag{3.3.121}$$

We have thus shown that

$$\begin{aligned} & \mathbb{E}\left[\|g-\widehat{g}_{\widehat{m}}\|^2_{\mathrm{L}^2}\, 1\left(\{\widehat{m}\leq m^*\}\right)\right] \\ \leq \quad & 3C_{\varkappa}\frac{C'^2_{\varphi}}{C^2_{\varphi}} \inf_{m\in\mathcal{M}}\left\{\|g-g_m\|^2_{\mathrm{L}^2}+\mathrm{pen}(m)\right\}, \end{aligned} \tag{3.3.122}$$

which is the desired result for the set $\{\widehat{m}\leq m^*\}$.

Next, consider the set $\{\widehat{m}>m^*\}$. We can decompose

$$\begin{aligned} & \|g-\widehat{g}_{\widehat{m}}\|^2_{\mathrm{L}^2} 1\left(\{\widehat{m}>m^*\}\right) \\ \leq \quad & 2\|g-g_{\widehat{m}}\|^2_{\mathrm{L}^2} 1\left(\{\widehat{m}>m^*\}\right)+2\|g_{\widehat{m}}-\widehat{g}_{\widehat{m}}\|^2_{\mathrm{L}^2} 1\left(\{\widehat{m}>m^*\}\right). \end{aligned} \tag{3.3.123}$$

First, we trivially have

$$\|g-g_{\widehat{m}}\|^2_{\mathrm{L}^2} 1\left(\{\widehat{m}>m^*)\}\right) \leq \|g-g_{m^*}\|^2_{\mathrm{L}^2}. \tag{3.3.124}$$

Next, we can estimate

$$\begin{aligned} & \|g_{\widehat{m}}-\widehat{g}_{\widehat{m}}\|^2_{\mathrm{L}^2} 1\left(\{\widehat{m}>m^*\}\right) \\ \leq \quad & \sum_{\substack{k>m^* \\ k\in\mathcal{M}}}\left\{\|g_{k'}-\widehat{g}_k\|-\frac{1}{2}\widetilde{\mathrm{pen}}(k)\right\}_+ \\ + \quad & \sum_{\substack{k>m^* \\ k\in\mathcal{M}}}\frac{1}{2}\widetilde{\mathrm{pen}}(k) 1\left(\{\widehat{m}=k\}\right)+\sum_{\substack{k>m^* \\ k\in\mathcal{M}}}\|g_k-g_{k'}\|^2 1\left(\{\widehat{m}=k\}\right). \end{aligned} \tag{3.3.125}$$

The expected value of the expression appearing in the second line of formula (3.3.125) is negligible by Proposition 3.3.15. Next, we find that on the set $\{\widehat{m}=k\}$,

$$\begin{aligned} \frac{1}{2}\widetilde{\mathrm{pen}}(k) \quad \leq \quad & -\|\widehat{g}_{m^*}\|^2_{\mathrm{L}^2}+\|\widehat{g}_k\|^2_{\mathrm{L}^2}-\frac{1}{2}\widetilde{\mathrm{pen}}(k)+\widetilde{\mathrm{pen}}(m^*) \\ \leq \quad & \left\{\|(g_{k'}-g_{m^{*\prime}})-(\widehat{g}_k-\widehat{g}_{m^*})\|^2_{\mathrm{L}^2}-\frac{1}{2}\widetilde{\mathrm{H}}^2(m^*,k)\right\}_+ \\ + \quad & \frac{1}{2}\widetilde{\mathrm{pen}}(m^*)+\|g_{k'}-g_{m^{*\prime}}\|^2_{\mathrm{L}^2} \end{aligned} \tag{3.3.126}$$

Once again, we have by Proposition 3.3.15, that

$$\mathbb{E}\left[\sum_{\substack{k>m^* \\ k\in\mathcal{M}}}\left\{\|(g_{k'}-g_{m^{*\prime}})-(\widehat{g}_k-\widehat{g}_{m^*})\|^2_{\mathrm{L}^2}-\frac{1}{2}\widetilde{\mathrm{H}}^2(m^*,k)\right\}_+\right] \leq Cn^{-1} \tag{3.3.127}$$

and again, we can estimate

$$\mathbb{E}\left[\widetilde{\mathrm{pen}}(m^*)\right] \leq \frac{C_\varphi'^2}{C_\varphi^2} C_{\mathrm{NK}}\,\mathrm{pen}(m^*). \tag{3.3.128}$$

Finally, we have by definition of $\widehat{m}$ on the set $\{\widehat{m} = k > m^*\}$:

$$\|g_{k'} - g_{m^{*\prime}}\|_{\mathrm{L}^2}^2 = \|g_{k'} - g_{m^*}\|_{\mathrm{L}^2}^2 \leq \|g - g_{m^*}\|_{\mathrm{L}^2}^2, \tag{3.3.129}$$

as well as

$$\|g_k - g_{k'}\|_{\mathrm{L}^2}^2 \leq \|g - g_{m^*}\|_{\mathrm{L}^2}^2. \tag{3.3.130}$$

Putting the above results together, we have shown that for positive constants C_{NK} and C depending only on the choice of the constants,

$$\begin{aligned} &\mathbb{E}\left[\|g - g_{\widehat{m}}\|_{\mathrm{L}^2}^2 \mathbb{1}\left(\{\widehat{m} > m^*\}\right)\right] \\ \leq\ & \frac{C_\varphi'^2}{C_\varphi} C_{\mathrm{NK}} \inf_{m \in \mathcal{M}} \left\{\|g - g_m\|_{\mathrm{L}^2}^2 + \mathrm{pen}(m)\right\} + Cn^{-1}, \end{aligned} \tag{3.3.131}$$

which is the desired result for the set $\{\widehat{m} > m^*\}$. This completes the proof. $\square$

Appendix A

Tools from Fourier analysis

We collect here some basic definitions and facts from Fourier analysis which are used throughout the text.

Definition A.1. *Let $g \in \mathrm{L}^1(\mathbb{R}^d)$. Then for $u \in \mathbb{R}^d$, the* Fourier transform *of g is defined to be*

$$\mathcal{F}g(u) := \int e^{i\langle u,x\rangle} g(x)\, dx \tag{A.1}$$

If $\mathbb{P}$ is a probability measure on $\mathbb{R}^d$, the Fourier transform of $\mathbb{P}$ is

$$\mathcal{F}\,\mathbb{P}(u) = \int e^{i\langle u,x\rangle}\, \mathbb{P}(\,dx). \tag{A.2}$$

The Fourier transform of a probability measure is often referred to as the characteristic function *of $\mathbb{P}$. For a random variable Z taking values in $\mathbb{R}^d$, $\mathcal{F}\,\mathbb{P}^Z$ is often denoted by φ_Z and referred to as the* characteristic function *of Z.*

Let us briefly summarise here the most important properties of the Fourier transformation.

Theorem A.2. *The following is true for the Fourier transformation of integrable functions, as well as for the Fourier transformation of probability measures:*

(i) Linearity:

$$\mathcal{F}(af + bg) = a\mathcal{F}f + b\mathcal{F}g. \tag{A.3}$$

(ii) Boundedness:

$$\|\mathcal{F}f\|_\infty \le \|f\|_{\mathrm{L}^1}, \quad \|\mathcal{F}\,\mathbb{P}\,\|_\infty \le 1. \tag{A.4}$$

(iii) For $z \in \mathbb{C}$, let $\bar{z}$ denote the complex conjugate. Then we have

$$\mathcal{F}f(-u) = \overline{\mathcal{F}f(u)}. \tag{A.5}$$

The following result shows the equivalence between differentiation and convolution operation and algebraic manipulations of the Fourier transform.

Theorem A.3.

(i) For $f, g \in \mathrm{L}^1(\mathbb{R}^d)$, the convolution of f and g is given by

$$(f * g)\,(x) := \int f(x-y)g(y)\, dy. \tag{A.6}$$

For finite Borel measures $\mathbb{P}$ and $\mathbb{Q}$ on $\mathbb{R}^d$, we have

$$\forall A \in \mathcal{B}(\mathbb{R}^d): \ (\mathbb{P} * \mathbb{Q})(A) := \int \mathbb{P}(A-y)\,\mathbb{Q}(\,dy). \tag{A.7}$$

The convolution operation is equivalent to pointwise multiplication in the Fourier domain:

$$\mathcal{F}(f*g)(u) = \mathcal{F}f(u)\mathcal{F}g(u) \quad \textit{and} \quad \mathcal{F}(\mathbb{P}*\mathbb{Q})(u) = \mathcal{F}\,\mathbb{P}(u)\mathcal{F}\,\mathbb{Q}(u). \tag{A.8}$$

(ii) Let $g \in C^k(\mathbb{R}^d)$ and assume that for every multi-index α with $|\alpha| \leq k$, $\partial^\alpha g \in \mathrm{L}^1(\mathbb{R}^d)$. Then we have

$$\mathcal{F}(\partial^\alpha g)(u) = (-iu)^\alpha \mathcal{F}g(u). \tag{A.9}$$

The proof of Theorem A.3 can be found, for example, in Chapter 7 in Rudin [62]. The next result is known as the *Fourier inversion formula.*

Theorem A.4. *Let $g \in \mathrm{L}^1(\mathbb{R}^d)$ and let $\mathcal{F}g \in \mathrm{L}^1(\mathbb{R}^d)$. Then we have for every continuity point x of g,*

$$g(x) = \frac{1}{(2\pi)^d}\int e^{-i\langle u,x\rangle}\mathcal{F}g(u)\,du. \tag{A.10}$$

Again, we refer to Chapter 7 in [62] for the proof.

Theorem A.5 (Riemann-Lebesgue-Lemma:)**.** *Let $g \in \mathrm{L}^1(\mathbb{R}^d)$ then the following holds:*

$$|\mathcal{F}g(u)| \to 0, \ |u| \to \infty. \tag{A.11}$$

The proof can be found, for example, in Chapter VI. 1 in Katznelson [39].

For functions in $\mathrm{L}^2(\mathbb{R}^d) \setminus \mathrm{L}^1(\mathbb{R}^d)$ the integral appearing in (A.1) is not well defined. However, the Fourier transformation can be given sense for arbitrary square integrable functions. This is done by considering the Fourier transformation on an appropriate subspace of the $\mathrm{L}^2(\mathbb{R}^d)$ (the Schwartz space), by noticing that the Fourier transformation is (up to a constant factor) an isometric isomorphism from the Schwartz space in itself and that the Schwartz space lies dense in $\mathrm{L}^2(\mathbb{R})$. This allows to continuously extend the Fourier transformation to the whole space $\mathrm{L}^2(\mathbb{R})$.

Theorem and Definition A.1. *There is a bijective linear mapping*

$$\widetilde{\mathcal{F}}: \mathrm{L}^2(\mathbb{R}^d) \to \mathrm{L}^2(\mathbb{R}^d) \tag{A.12}$$

such that the following holds true:

(i) The restriction of $\widetilde{\mathcal{F}}$ to $\mathrm{L}^2(\mathbb{R}^d) \cap \mathrm{L}^1(\mathbb{R}^d)$ agrees with $\mathcal{F}$ as defined in Definition A.1.

(ii) For arbitrary f in $\mathrm{L}^2(\mathbb{R}^d)$, we have

$$\|f\|_{\mathrm{L}^2}^2 = \frac{1}{(2\pi)^d}\|\mathcal{F}f\|_{\mathrm{L}^2}^2. \tag{A.13}$$

We call $\widetilde{\mathcal{F}}f$ *the* Fourier transformation *on* $\mathrm{L}^2(\mathbb{R})$ *and simply write* $\mathcal{F}$ *instead of* $\widetilde{\mathcal{F}}$.

For the proof of the statement, we refer, to chapter 7 in [62].

Theorem A.6. *Let* $f, g \in \mathrm{L}^2(\mathbb{R})$. *Then we have*

$$\langle f, g \rangle = \frac{1}{(2\pi)^d} \langle \mathcal{F}f, \mathcal{F}g \rangle. \tag{A.14}$$

This formula is known as the Plancherel formula.

Appendix B

Distributions

We give here a short overview of the theory of distributions and introduce the notation which is used throughout the text. For sake of simplicity, we content ourselves with considering the one dimensional case.

Definition B.1. *Let $\Omega \subseteq \mathbb{R}$ be a nonempty open set. For a compact set $\mathrm{K} \subseteq \Omega$, let*

$$\mathcal{D}_{\mathrm{K}}(\Omega) := \{\varphi : \mathbb{R} \to \mathbb{C} \,|\, \operatorname{supp}(\varphi) \subseteq \mathrm{K}, \varphi \in C^\infty(\Omega)\} \tag{B.1}$$

be the space of infinitely differentiable functions with support in K.

Moreover, let

$$\mathcal{D}(\Omega) := \bigcup_{\mathrm{K}} \mathcal{D}_{\mathrm{K}}(\Omega) \tag{B.2}$$

be the space of infinitely differentiable functions on Ω which are compactly supported.

The functions collected in $\mathcal{D}(\Omega)$ are called test functions.

Definition B.2. *On $\mathcal{D}_{\mathrm{K}}(\Omega)$, a system of semi-norms is defined as follows: For $k \in \mathbb{N}$,*

$$p_{\mathrm{K},k}(\varphi) := \sup_{m \le k} \sup_{x \in \mathrm{K}} \left| \varphi^{(m)}(x) \right|. \tag{B.3}$$

To avoid a general discussion about topological concepts, we do not introduce the locally convex topology on $\mathcal{D}(\Omega)$ and define, in an abstract way, distributions as linear functionals which are continuous with respect to this topology, see, for example, Definition 6.7 in Rudin [62]. Instead, using Theorem 14.4 in Jantscher [35], we work with the following definition, which is more intuitive:

Definition B.3. *A linear mapping $f : \mathcal{D}(\Omega) \to \mathbb{C}$ is called as* distribution, *if for every compact set* K, *there is a positive constant c and a constant $k \in \mathbb{N}$ such that for every $\varphi \in \mathcal{D}_{\mathrm{K}}(\Omega)$,*

$$|f(\varphi)| \le c\, p_{\mathrm{K},k}(\varphi). \tag{B.4}$$

Every locally integrable function can be identified, in a natural way, with a distribution:

Theorem and Definition B.1. *Let $\widetilde{f}$ be a measurable and locally integrable function, that is, for every* K, $\int_{\mathrm{K}} |\widetilde{f}|(x)\, dx < \infty$. *Then the linear mapping*

$$f : \mathcal{D}(\Omega) \to \mathbb{C},\ \varphi \mapsto \int \varphi(x) \widetilde{f}(x)\, dx \tag{B.5}$$

is a distribution.

A distribution f is called regular *if there is a locally integrable function $\widetilde{f}$ such that f is given by (B.5).*

A distribution is called singular *if it is not regular.*

The proof of this statement can be found, for example, in Chapter 6 in [62].

If there is no danger of confusion, we often identify the regular distribution f with the function $\widetilde{f}$. Moreover, we use, even for non-regular distributions f, the notation $\int f(x)g(x)\,\mathrm{d}x$ instead of $f(g)$.

Definition B.4. *Let $V \subseteq \Omega$ be an open set. Then a distribution f* vanishes on V, *if $\int \varphi(x)f(x)\,dx = 0$ holds for every test function φ with support in V. The* support *of a distribution is the complement of the union of all open sets on which f vanishes.*

Definition B.5. *Let f be a distribution and* K *some compact set. The* order of f on K *is the smallest positive integer k for which a constant $c > 0$ exists such that (B.4) holds.*

A distribution is called of finite order *if there is a constant k_0 such that for every compact set* K*, the order of f on* K *is smaller than or equal to k_0. In this case, k_0 is called the* order of f.

Lemma B.6. *If f has compact support, then f is of finite order. In this case, there exists a constant $c < \infty$ and a nonnegative integer k such that the following holds true for every compact set* K *and for every test function φ:*

$$f(\varphi) \leq c\,p_{K,k}(\varphi). \tag{B.6}$$

The proof can be found in [62], Theorem 6.24.

Definition B.7. *In the situation of the preceding Lemma, we let $\|f\|$ denote the smallest nonnegative constant for which (B.6) holds true.*

Definition B.8. *The* Schwartz space $\mathcal{S}(\mathbb{R})$ *consists of all functions $f \in C^\infty(\mathbb{R})$ for which the following holds:*

$$\forall m \in \mathbb{N}\,\forall \alpha \in \mathbb{N} : \lim_{|x|\to\infty} x^\alpha f^{(m)}(x) = 0. \tag{B.7}$$

The Schwartz space is endowed with the system of semi-norms

$$p_{j,k}(\varphi) = \sup_{x\in\mathbb{R}} \sup_{m\leq k} (1 + |x|^j)|\varphi^{(m)}(x)|. \tag{B.8}$$

Lemma 2.0.16. *The Fourier transformation is, up to a constant factor, an isometric isomorphism of the Schwartz space onto itself.*

Definition B.9. *A linear mapping $f : \mathcal{S}(\mathbb{R}) \to \mathbb{C}$ is called a* tempered distribution, *if there is a positive constant a and $j, k \in \mathbb{N}$ such that for any $\varphi \in \mathcal{S}$:*

$$|f(\varphi)| \leq a p_{j,k}(\varphi). \tag{B.9}$$

The space of tempered distributions is denoted by $\mathcal{S}'(\mathbb{R})$.

Again, we often use the notation $\int \varphi(x)f(x)\,\mathrm{d}x$ instead of $f(\varphi)$.

The restriction of a tempered distribution to $\mathcal{D}(\mathbb{R})$ is a distribution, see Theorem 37.1 in [35], but the converse is false: A distribution need not be tempered, so the space $\mathcal{S}'(\mathbb{R})$ of tempered distributions is a proper subspace of the space $\mathcal{D}'(\mathbb{R})$ of distributions.

The Fourier transformation can be given sense not only for integrable and square integrable functions, but also for the much broader class of tempered distributions, see for example, Chapter 9 in [35]:

Theorem and Definition B.2. *Let f be a tempered distribution. Then there is a tempered distribution $\widetilde{\mathcal{F}}f$ such that the following equality holds:*

$$\forall \varphi \in \mathcal{S}(\mathbb{R}): \ \frac{1}{2\pi}\int \widetilde{\mathcal{F}}f(u)\mathcal{F}\varphi(-u)\,du = \int f(x)\varphi(x)\,dx. \tag{B.10}$$

$\widetilde{F}$ is called the Fourier transform *of f. We write, henceforth $\mathcal{F}f$ instead of $\widetilde{\mathcal{F}}f$. The Fourier transformation $\mathcal{F}: \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ thus defined is a bijective linear mapping of the dual space of the Schwartz space onto itself.*

Formula B.10 is known as Parseval's identity.

It is important to note that the above definition is consistent with the usual notion of the Fourier transform. Again, we refer to Chapter 9 in [35] for the proof of the following statement:

Theorem B.10. *Let μ be a finite Borel measure.Then formula (B.10) is satisfied for $\mathcal{F}\mu(u) = \int e^{iux}\mu(\,dx)$. Let $f \in \mathrm{L}^1(\mathbb{R}) \cup \mathrm{L}^2(\mathbb{R})$. Let $\mathcal{F}f$ be defined as in the preceding section. Then formula (B.10) is satisfied for $\mathcal{F}f$.*

We will have to extend the theory of linear functionals from the test function space to more general function spaces.

Definition B.11. *Let $\Omega \subseteq \mathbb{R}$ be an open set. Let $C^k(\Omega)$ denote the space of k-times continuously differentiable functions on Ω. Let $C^k(\Omega)$ be endowed with the following topology: For a sequence (φ_n) in $C^k(\Omega)$, we say that $\varphi_n \underset{k}{\Rightarrow} 0$, if for every compact set $\mathrm{K} \subseteq \Omega$ and every $m \leq k$,*

$$\sup_{x\in\mathrm{K}} \left|\varphi^{(m)}(x)\right| \to 0. \tag{B.11}$$

Then one can show the following statement:

Theorem B.12. *Every distribution f of order k with compact support in Ω can be extended, in a unique way, to a continuous linear functional on $C^k(\Omega)$. Conversely, the restriction of a continuous linear functional on $C^k(\Omega)$ to $\mathcal{D}(\Omega)$ is a distribution of order k.*

It is shown in §16 in Jantscher [35] that every distribution of order k in $\mathcal{D}(\Omega)$ can be extended in a unique way to a continuous linear functional on the space $\mathcal{D}^k_{\mathrm{K}}(\Omega)$ of k-times continuously differentiable functions with compact support in Ω. On the other hand, it is shown in §30 therein that every distribution having compact support in Ω can be extended to a linear functional on the space of infinitely differentiable functions $C^\infty(\Omega)$. A straightforward combination of both statements gives the desired result.

We will often use the following extension of Parseval's identity:

Theorem B.13. *Let μ be a finite Borel measure. Let $f \in \mathrm{L}^1(\mathbb{R})$ and $\sup_{x\in\mathbb{R}} |f(x)| < \infty$. Assume that $\mathcal{F}f\mathcal{F}\mu(-\bullet) \in \mathrm{L}^1(\mathbb{R})$. Then we have*

$$\int f(x)\mu(\,dx) = \frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\mu(u)\,du. \tag{B.12}$$

Let μ be a finite Borel measure. Let f be a distribution with compact support K *and order k. Assume that for some open set, the restriction $\mu\big|_D$ has a Lebesgue density $g_D \in C^k(D)$. Let $\int(x)\mu(\,dx) := \int f(x)g_D(x)\,dx$, which makes sense in view of the preceding theorem. Assume that $\mathcal{F}f(-\bullet)\mathcal{F}\mu \in \mathrm{L}^1(\mathbb{R})$. Then*

$$\int f(x)\mu(\,dx) = \frac{1}{2\pi}\int \mathcal{F}f(-u)\mathcal{F}\mu(u)\,du \tag{B.13}$$

holds true.

Proof (sketch). The proof relies on the fact that any distribution can be approximated, with arbitrary precision, by regular distributions, which can be identified with Schwartz functions, see Theorem 39.5 in [35]. Moreover, the approximating functions can be chosen such that their Fourier transform has compact support. This will permit to use Parseval's identity (B.10) and dominated convergence to find that for a sequence $f_k \to f$ to be appropriately chosen,

$$\begin{aligned}\int f(x)\mu(\mathrm{d}x) &= \lim_{k\to\infty}\int f_k(x)\mu(\mathrm{d}x)\\ &= \lim_{k\to\infty}\int \mathcal{F}f_k(-u)\mathcal{F}\mu(u)\,\mathrm{d}u = \int \mathcal{F}f(-u)\mathcal{F}g(u)\,\mathrm{d}u.\end{aligned} \tag{B.14}$$

□

Finally, we need the following important result which is known as Paley-Wiener Theorem:

Theorem B.14.

(i) If f is a distribution with support in $(-r, r)$ and has order k, then $\mathcal{F}f$ is an entire function and we have

$$\forall z \in \mathbb{C}: \; |\mathcal{F}f(z)| \leq \gamma(1+|z|)^k e^{r|\operatorname{Im} z|}. \tag{B.15}$$

for some constant $\gamma < \infty$.

(ii) Conversely, if g is an entire function such that

$$\forall z \in \mathbb{C}: \; |g(z)| \leq \gamma(1+|z|)^k e^{r|\operatorname{Im} z|} \tag{B.16}$$

holds for some constant $\gamma < \infty$, then there exists a distribution f of order k with support in $(-r, r)$ such that for any $u \in \mathbb{R}$, $g(u) = \mathcal{F}f(u)$.

For the proof of this statement, see Theorem 7.23 in [62].

Appendix C

The general case

We conclude by giving a very short discussion on the general case, where X is no longer assumed to have finite variation on compact sets, a drift part is admitted and there may be a non-zero Gaussian component. We only formulate the following assumption:

Assumption C.1. *X has finite fourth moments, so*

$$\int x^4 \nu(\,dx) < \infty. \tag{C.1}$$

The model

We observe

$$X_\Delta, \cdots, X_{2n\Delta}, \tag{C.2}$$

where X is a Lévy process for which Assumption C.1 is satisfied. In this case, $\mu(\,\mathrm{d}x) = x\nu(\,\mathrm{d}x)$ need no longer be finite, but $\mu_\sigma(\,\mathrm{d}x) := \sigma^2\delta_0(\,\mathrm{d}x) + x^2\nu(\,\mathrm{d}x)$ is a finite measure and we may follow Neumann and Reiß and consider the alternative parametrisation given in part (i) of Lemma 1.1.11. In a high frequency framework, the general case has been treated by Comte and Genon-Catalot [19].

The center γ_1 equals the mean value of X_1, which allows to use $\widehat{\gamma}_1 = \frac{1}{T}X_T$ as an estimator of γ_1.

The Fourier transform of μ can be recovered, using the first and second derivative of the characteristic function. We have in this case

$$\begin{aligned} \mathcal{F}\mu(u) &= \int e^{iux}\mu(\,\mathrm{d}x) = -\int \frac{\partial^2}{\partial u^2}\frac{e^{iux} - 1 - iux}{x^2}\mu(\,\mathrm{d}x) \\ &= -\Psi''(u) = \frac{1}{\Delta}\left(\frac{(\varphi'_\Delta(u))^2}{(\varphi_\Delta(u))^2} - \frac{\varphi''_\Delta(u)}{\varphi_\Delta(u)}\right). \end{aligned} \tag{C.3}$$

The estimator

Suppose that we want to estimate a linear functional of μ:

$$\theta = \int f(x)\mu(\,\mathrm{d}x). \tag{C.4}$$

We impose the same regularity conditions on μ and on f as previously. In the usual way, we define the empirical versions $\widehat{\varphi}_{\Delta,n}$ of φ_Δ, $\widehat{\varphi}'_{\Delta,n}$ of φ'_Δ and $\widehat{\varphi}''_{\Delta,n}$ of

φ''_Δ, as well as the estimator $\frac{1}{\widetilde{\varphi}_{\Delta,n}}$ of $\frac{1}{\varphi_\Delta}$. In analogy with Definition 2.2.7, we can then define a kernel estimator of μ.

Definition 3.0.17. *For a kernel function* K *satisfying (K1) and (K2) and a bandwidth $h > 0$, let the corresponding kernel estimator of θ be*

$$\widehat{\theta}_{\Delta,h,n} = \int \mathcal{F}f(-u)\mathcal{F}\mathrm{K}_h(u)\frac{1}{\Delta}\left(\frac{\left(\widehat{\varphi}'_{\Delta,n}(u)\right)^2}{(\widetilde{\varphi}_{\Delta,n}(u))^2} - \frac{\widehat{\varphi}''_{\Delta,n}(u)}{\widetilde{\varphi}_{\Delta,n}(u)}\right) du. \tag{C.5}$$

Discussion

Let us formulate some ideas about what might happen in the general case:

One can formulate the following conjecture about the risk bounds:

Conjecture C.2. *Let $\widehat{\theta}_{\Delta,h,n}$ be the kernel estimator introduced in Definition 3.0.17. Then we obtain the following bound on the squared risk:*

$$\begin{aligned}
&\mathbb{E}\left[|\theta - \widehat{\theta}_{\Delta,h,n}|^2\right] \\
\leq\ & \left|\int f(x)\mu(\,dx) - \int f(x)(\mathrm{K}_h * \mu)(\,dx)\right|^2 \\
+\ & \frac{T^{-1}}{2\pi^2}\left\{C_1 \int |\mathcal{F}\mathrm{K}_h(u)|^2 \frac{|\mathcal{F}f(-u)|^2}{|\varphi_\Delta(u)|^2}\, du \wedge C_2 \left(\int |\mathcal{F}\mathrm{K}_h(u)|\, \frac{|\mathcal{F}f(-u)|}{|\varphi_\Delta(u)|}\, du\right)^2,\right\}
\end{aligned}$$

with constants C_1 and C_2 depending on $\Psi^{(k)},\ k = 1, \cdots, 4$.

The proof will essentially rely on an extension of Lemma 2.5.1. We will have to consider

$$\frac{1}{\Delta}\,\mathbb{E}\left[T(u)T(-v)\right], \tag{C.6}$$

with T defined by

$$T(x) := \left(\frac{\widehat{\varphi}''_{\Delta,n}(x)}{\widetilde{\varphi}_{\Delta,n}(x)} - \frac{\varphi''_\Delta(x)}{\varphi_\Delta(x)}\right) - \left(\frac{\left(\widehat{\varphi}'_{\Delta,n}(x)\right)^2}{(\widetilde{\varphi}_{\Delta,n}(x))^2} - \frac{(\varphi'(x))^2}{(\varphi(x))^2}\right). \tag{C.7}$$

We will obtain a decomposition, similar to what happens in the proof of Lemma 2.5.1 and the arguments will basically run along the same lines, but with the double number of terms. There will appear covariance terms of the form

$$\mathrm{Cov}(\widehat{\varphi}''_1(u), \widehat{\varphi}''_1(v)) = \varphi^{(iv)}(u - v) - \varphi''(u)\varphi''(-v),$$

which lead to the occurrence of $\Psi^{(k)}, k = 1, \cdots, 4$.

If one can prove the Conjecture C.2, the rate results presented in Section 2.3 can immediately be extended to the case of infinite variation and nonzero Gaussian part.

Moreover, we believe that the model selection procedure can be extended to the general case and one can prove results in analogy with Theorem 3.2.11 and Theorem 3.2.12.

The uniform control on the characteristic function in the denominator, which is essential for our arguments, does not depend on the assumption that the Gaussian part and drift is zero and the jumps are moderate. The extension of Proposition 3.3.10 might cause some technical difficulties when dealing with the squared terms. Possibly, there will arise the necessity to introduce another sample splitting and define two independent estimators $\widehat{\varphi}'_{\Delta,n,1}$ and $\widehat{\varphi}'_{\Delta,n,2}$ of φ'_Δ and then use $\widehat{\varphi}'_{\Delta,n,1}\widehat{\varphi}'_{\Delta,n,2}$ as an unbiased estimator of $(\varphi'_\Delta)^2$.

All in all, we believe that the loss of generality which results from only treating the special case in this thesis is small, compared to the loss of intuition and readability which would have resulted from dropping the assumptions *(A1)-(A4)*. So the decision to treat only the special case can be seen as a reasonable bias-variance-compromise.

Bibliography

[1] Hirotugu Akaike. Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory*, pages 267–281, 1973.

[2] David Applebaum. *Lévy processes and stochastic calculus*. Cambridge University Press, 1 edition, 2004.

[3] Denis Belomestny. Spectral estimation of the fractional order of a Lévy process. *The Annals of Statistics*, 38(1):317–351, 2010.

[4] Denis Belomestny and Markus Reiß. Spectral calibration of exponential Lévy models. *Finance and Stochastics*, 10(4):449–474, 2006.

[5] Lucien Birgé. An alternative point of view on Lepski's method. In *State of the art in probability and statistics: Festschrift for Willem R. van Zwet*, pages 113–133. IMS Lecture Notes-Monograph Series, 1999.

[6] Lucien Birgé and Pascal Massart. From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, 1997.

[7] Olivier Bousquet. Concentration inequalities for sub-additive functions using the entropie method. *Stochastic inequalities and applications*, 56:213–247, 2003.

[8] Christina Butucea. Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics*, 32(2):181–192, 2004.

[9] Christina Butucea and Fabienne Comte. Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli*, 15(1):69–98, 2009.

[10] T. Tony Cai and Mark G. Low. Adaptive estimation of linear functionals under different performance measures. *Bernoulli*, 11(2):341–358, 2005.

[11] T. Tony Cai and Mark G. Low. On adaptive estimation of linear functionals. *Annals of Statistics*, 33(5):2311–2343, 2005.

[12] Peter Carr, Hélyette Geman, Dilip Madan, and Marc Yor. The fine structure of asset returns: An empirical investigation. *Journal of Business*, 75(2):305–332, 2002.

[13] Peter Carr, Hélyette Geman, Dilip Madan, and Marc Yor. Stochastic volatility for Lévy processes. *Mathematical Finance*, 13(3):345–382, 2003.

[14] Peter Carr, Dilip Madan, and Eric Chang. The variance Gamma process and option pricing. *European Finance Review*, 2:79–105, 1998.

[15] Peter Carr and Liuren Wu. Time changed Lévy processes and option pricing. *Journal of Financial Economics*, 71:113–141, 2004.

[16] Raymond J. Carroll and Peter Hall. Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186, 1988.

[17] Fabienne Comte and Valentine Genon-Catalot. Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Processes and their Applications*, 119:4088–4123, 2009.

[18] Fabienne Comte and Valentine Genon-Catalot. Nonparametric adaptive estimation for pure jump Lévy processes. *Annales de l'Institut Henri Poincaré*, 3(46):595–617, 2010.

[19] Fabienne Comte and Valentine Genon-Catalot. Estimation for Lévy processes from high frequency data within a long time interval. *Annals of Statistics*, 39(2):803–837, 2011.

[20] Fabienne Comte and Claire Lacour. Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B*, 73:601–627, 2011.

[21] Fabienne Comte, Yves Rosenholc, and Marie-Luce Taupin. Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, (34):431–452, 2006.

[22] Rama Cont and Peter Tankov. *Financial modelling with jump processes.* Chapman & Hall/CRC Financial Mathematics Series. Boca Raton, 2004.

[23] Richard M. Dudley. *Uniform central limit theorems.* Cambridge Studies in Advanced Mathematics 63. Cambridge: Cambridge University Press, 2008.

[24] Ernst Eberlein and Ulrich Keller. Hyperbolic distributions in finance. *Bernoulli*, 1:281–299, 1995.

[25] Sam Efromovich. Density estimation for the case of supersmooth measurement errors. *Journal of the American Statistical Association*, 92:526–535, 1997.

[26] Jianqing Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272, 1991.

[27] José Figueroa-Lopez and Christian Houdré. Small-time expansions for the transition distribution of Levy processes. *Stochastic Processes and their Applications*, 119:3862–3889, 2009.

[28] José E. Figueroa-López. Small-time moment asymptotics for Lévy processes. *Statistics and Probability Letters*, 78(18):3355–3365, 2008.

[29] José E. Figueroa-López. Nonparametric estimation for Lévy models based on discrete sampling. *IMS Lecture Notes of the 3rd E.L. Lehmann Symposium*, 57:117–146, 2009.

[30] José E. Figueroa-López and Christian Houdré. Risk bounds for the non-parametric estimation of Lévy processes. *IMS Lecture Notes - Monograph Series; High Dimensional Probability IV*, 51:96–116, 2006.

[31] Alexander Goldenshluger and Sergei V. Pereverzev. On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli*, 9(5):783–807, 2003.

[32] Alexander Goldshluger. On pointwise adaptive nonparametric deconvolution. *Bernoulli*, (5):907–925, 1999.

[33] Shota Gugushvili. Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *Journal of Nonparametric Statistics*, 21(3):321–343, 2009.

[34] Shota Gugushvili. Nonparametric estimation for discretely sampled Lévy processes. *Annales de l'Institut Henri Poincaré*, 48(1):282–307, 2012.

[35] Lothar Jantscher. *Distributionen*. De Gruyter, 1971.

[36] Jan Johannes. Deconvolutionwith unknown error distribution. *The Annals of Statistics*, 37(5a):2301–2323, 2009.

[37] Geurt Jongbloed, Frank H. van der Meulen, and Aad W. van der Vaart. Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 11(5):759–791, 2005.

[38] Johanna Kappus and Markus Reiß. Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Statistica Neerlandica*, 64:314–328, 2010.

[39] Yitzhak Katznelson. *An introduction to harmonic analyis*. Cambridge University Press, 3 edition, 2004.

[40] Ulrich Keller. *Realistic modelling of financial derivatives*. PhD thesis, Universität Freiburg, 1997.

[41] Thierry Klein. Une inégalité de concentration á gauche pour les processus empiriques. *Comtes Rendus Mathématique*, 334:501–504, 2002.

[42] Thierry Klein and Emmanuel Rio. Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060–1077, 2005.

[43] Uwe Küchler and Stefan Tappe. On the shape of bilateral gamma densities. *Statistics and Probability Letters*, 78:2478–2484, 2008.

[44] Béatrice Laurent, Carenne Ludeña and Clémentine Prieur. Adaptive estimation of linear functionals by model selection. *Electronic Journal of Statistics*, 2:993–1020, 2008.

[45] Michel Ledoux. On talagrand's deviation inequalities for product measures. *ESAIM: Probability and statistics*, 1:63–87, 1996.

[46] Oleg V. Lepski. On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and Applications*, 35:454–466, 1990.

[47] Oleg V. Lepski. Asymptotically minimax adaptive estimation. I.Upper bounds. *Theory of Probability and Applications*, 36(3):682–697, 1991.

[48] Oleg V. Lepski. Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. *Theory of Probability and Applications*, 37(3):433–448, 1992.

[49] Oleg V. Lepski and Vladimir Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics*, 25(6):2512–2546, 1997.

[50] Colin L. Mallows. Some comments on C_p. *Technometrics*, 15(4):661–675, 1973.

[51] Pascal Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *The Annals of Probability*, 28:863–884, 2000.

[52] Pascal Massart. *Concentration Inequalities and Model selection*. Springer, 2003.

[53] Pascal Massart and Lucien Birgé. Gaussian Model selection. *Journal of the European Mathematical Society*, 3(3):203–268, 2001.

[54] Alexander Meister. *Deconvolution Problems in Nonparametric Statistics*. Number 193 in Lecture Notes in Statistics. Springer, 2009.

[55] Michael Neumann. Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis*, 98:1955–1968, 2006.

[56] Michael Neumann and Markus Reiß. Nonparametric estimation for Lévy processes from low frequency observations. *Bernoulli*, 15(1):223–248, 2009.

[57] Michael H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330, 1997.

[58] Richard Nickl and Markus Reiß. A Donsker Theorem for Lévy measures. Technical Report 3, SFB 649 Economic Risk, 2012.

[59] Steven Orey. On continuity properties of infinitely divisible functions. *The Annals of Mathematical Statistics*, 39(3):936–937, 1968.

[60] Emmanuel Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probability Theory and Related Fields*, 119:163–175, 2001.

[61] Emmanuel Rio. Une inégalité de bennett pour les maxima de processus empiriques. *Annales de l'Institut Henri Poincaré*, 38:1053–1057, 2002.

[62] Walter Rudin. *Functional analysis.* McGraw-Hill, 2 edition, 1991.

[63] Ken-Iti Sato. *Lévy processes and infinitely divisible distributions.* Cambridge University Press, 1999.

[64] Jakob Söhl. Confidence sets in nonparametric calibration of exponential Lévy models. Technical Report 12, SFB 649 Economic Risk, 2012.

[65] Leonard A. Stefanski. Rates of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters*, 9:229–235, 1990.

[66] Michel Talagrand. Sharper bounds for empirical processes. *Annals of Probability*, 22:28–76, 1994.

[67] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Instutut H.E.S.*, 81:73–205, 1996.

[68] Mathias Trabs. Calibration of self-decomposable Lévy models. Technical Report 73, SFB 649 Economic Risk, 2011.

[69] Alexandre B. Tsybakov. Pointwise and sup-norm sharp adaptive estimation of functionals on the Sobolev classes. *Annals of statistics*, (26):2420–2469, 1998.

[70] Alexandre B. Tsybakov. *Introduction à l' éstimation non-paramétrique.* Number 41 in Mathematiques et Applications. Springer, 2004.

[71] Alexandre B. Tsybakov and J. Klemelä. Sharp adaptive estimation of linear functinoals. *Annals of Statistics*, 29(6):1567–1600, 2001.

[72] Jeannette H.C. Woerner. *Statistical analysis for discretely observed Lévy processes.* PhD thesis, Albert-Ludwigs-Universität Freiburg, 2001.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Rostock, den 21.05.2012

Johanna Kappus